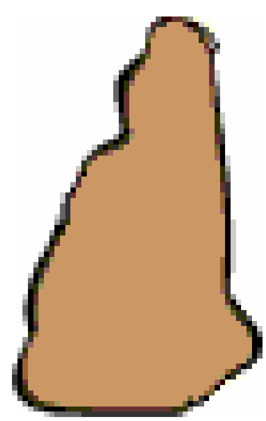
New Hampshire Department of Education

The Nations' Report Card
New Hampshire
Department of Education
2003 NAEP State Report
Grade 8 Mathematics



FORWARD

Although this report was put together in final form by New Hampshire Department of Education staff there are a number of other significant contributors who made its outcome possible.

First, we acknowledge the many schools' students and staff who gave of their time and energy to participate in the 2003 State National Assessment of Educational Progress (NAEP). As the New Hampshire sample they allowed an estimate of what grade four and grade eight students in our state and the nation know and can do in mathematics and reading. Without them of course there would be no data; nothing to report. The last time New Hampshire had data of this type was in 1998 so the 2003 State NAEP assessment was a significant event.

Equally as important is the work done by the National Center of Education Statistics and its contractors who systematically gathered, scored, and organized the results in usable tables and graphs. This work made the monumental task of ferreting out recognizable results manageable, providing valuable opportunities for analysis. We are in debt as well to the wonderful and helpful people at the NAEP State Service Center. They provided excellent training and support on a continual basis to assure the highest level of success in all the state NAEP endeavors.

Special recognition as well is given to the designers of the State Report Generator (SRG) that allowed customized state report generation from the voluminous data gathered. The selection and filtering mechanisms made possible the generation of this and other New Hampshire NAEP reports in a timely fashion. Nancy Mead and her colleagues at the Educational Testing Service gave us a superb product to utilize.

Finally, recognition and thanks is given to the many persons in the Bureau of Accountability at the New Hampshire Department of Education who provided guidance and expertise in shaping the final report products. A special "Thank You" is set aside for Carol Angowski whose creative and technical skill was essential in producing this and a number of New Hampshire NAEP-related published documents.

Contact Information
Testing Director: Tim Kurtz
NAEP Coordinator: David Gebhardt
NH State Dept. of Education

Table of Contents

Eighth Grade Mathematics Report

Overall Student, School/District Characteristics 2002-2003

Student Characteristics

Number enrolled: **207,671**
Percent in Title I schools: **48.7 %**
With Individualized Education Programs (IEP): **13.9%**
Percent in limited-English proficiency programs: **1.57%**
Percent eligible for free/reduced lunch: **16.39%**

Racial/Ethnic Background

White: **94.4%**
Black: **1.5%**
Hispanic: **2.2%**
Asian/Pacific Islander: **1.6%**
American Indian/Alaskan Native: **0.3%**

School/District Characteristics

Number of SAUs: **84**
Number of school districts: **176**
Number of schools: **466**
Number of charter schools: **N/A**
Per-pupil expenditures: **$7,233**[1]
Pupil/teacher ratio: **13.5**
Number of FTE teachers: **14,478**

'--' : data unavailable Source: Common Core of Data, *2002-2003* school year
[1] Common Core of Data, *2002-2003* school year

The Nations' Report Card
MATHEMATICS 2003

New Hampshire
Grade 8
Public Schools

NEW HAMPSHIRE NAEP STATE REPORT

KEY FINDINGS

For grade 8:

- The average mathematics scale score for students in New Hampshire was 286. This was higher than that of 1990 (273) and was higher than that in 1992 (278).
- New Hampshire's average score (286) was higher than that of the nation's public schools (276).
- Students' average scores in New Hampshire were higher than those in 40 jurisdictions, not significantly different from those in 11 jurisdictions, and lower than those in 1 jurisdiction.
- The percentage of students in New Hampshire who performed at or above the *Proficient* level was 35 percent. This was greater than that in 1990 (20 percent) and was greater than that in 1992 (25 percent).
- In New Hampshire, the percentage of students who performed at or above *Proficient* was higher than that for the nation's public schools (27 percent).

This report provides selected results from the National Assessment of Educational Progress (NAEP) for New Hampshire's public-school students at grade 8. Since 1990, mathematics has been assessed in five different years at the state level (at grade 8 in 1990, and at both grades 4 and 8 in 1992, 1996, 2000, and 2003). In 2003, 53 jurisdictions participated: the 50 states, District of Columbia, Department of Defense Domestic Dependent Elementary and Secondary Schools, and Department of Defense Dependents Schools (Overseas). New Hampshire participated and met the criteria for reporting public-school results at grade 8 in 1990, and at both grades 4 and 8 in 1992 and 2003.

NAEP is a project of the National Center for Education Statistics (NCES). For more information about the assessment, see *The Nation's Report Card, Mathematics Highlights 2003* or *The Nation's Report Card: Mathematics 2003*, which will be available in 2004. The full set of results is available in an interactive database on the NAEP web site (http://nces.ed.gov/nationsreportcard/). Released test questions, scoring guides, and question-level performance data are also available on the web site.

The U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP) has provided software that generated user-selectable data, statistical significance test result statements, and technical descriptions of the NAEP assessments for this report. Content may be added or edited by states or other jurisdictions. This document, therefore, is not an official publication of the National Center for Education Statistics.

Introduction

What Was Assessed?

The content for each NAEP assessment is determined by the National Assessment Governing Board (NAGB). The development process for mathematics required the active participation of teachers, curriculum specialists, subject-matter specialists, local school administrators, parents, and members of the general public. The objectives for each NAEP assessment are described in a "framework," a document that delineates the important content and process areas to be measured, as well as the types of questions to be included on the assessment.

The mathematics framework for the 2003 National Assessment of Educational Progress replicates the frameworks that guided the 1996 and 2000 mathematics assessments. This framework was developed under the auspices of the College Board and directed by NAGB. The framework calls for questions based on five mathematics content areas: number sense, properties and operations; measurement; geometry and spatial sense; data analysis, statistics and probability; and algebra and functions. Questions were also categorized according to two domains: mathematical abilities and mathematical power. Mathematical abilities describes the types of knowledge or processes required for a student to successfully respond to a question. Mathematical abilities may reflect conceptual understanding, procedural knowledge, or a combination of both in problem solving. The second domain, mathematical power, reflects the processes stressed as major goals of the mathematics curriculum. These include the student's ability to reason, to communicate, and to make connections between concepts and skills either across the mathematics content areas, or from mathematics to other curricular areas.

The framework also incorporates the use of calculators (four-function at grade 4 and scientific at grade 8), rulers, protractors (grade 8), and manipulatives such as spinners and geometric shapes. The use of these ancillary materials and the use of calculators were incorporated into some parts of the assessment, but not all. Calculator use was permitted on approximately one-third of the test questions. The mathematics framework is available on the NAGB web site (http://www.nagb.org/pubs/math_framework/toc.html).

A combination of multiple-choice and constructed-response questions was used to assess students' mathematics abilities. Short constructed-response questions ask students to provide the answer for a numerical problem or to briefly describe the solution to a problem. Longer constructed-response questions require students to produce both a solution and a short paragraph describing the solution or its interpretation. For a number of these questions, students can use calculators, protractors, or rulers. Released test questions, along with student performance data by state, are available on the NAEP web site (http://nces.ed.gov/nationsreportcard/itmrls/).

Who Was Assessed?

In 2003, 53 jurisdictions participated in NAEP: the 50 states, District of Columbia, Department of Defense Domestic Elementary and Secondary Schools, and Department of Defense Dependents Schools (Overseas). The target sample for each state or other jurisdiction was approximately 100 schools at a grade and approximately 3,000 students for each subject at a grade, except in small or sparsely populated jurisdictions. The sample of schools and students was chosen in a two-stage sampling process. First, the sample of schools was selected by probability sampling methods. Then, within the participating schools, random samples of students were chosen. Beginning in 2002, the national sample was obtained by aggregating the samples from each state. The national results include the results from the states, weighted appropriately to represent the U.S. student population. Only public schools, however, are included in the state reports. The overall participation rates for schools and students must meet guidelines established by the National Center for Education Statistics (NCES) and the National Assessment Governing Board (NAGB) in order for assessment results to be reported publicly. Data are not reported to the public for a state or jurisdiction that participates but does not meet minimum participation guidelines (see http://nces.ed.gov/nationsreportcard/about/participrates.asp). Participation rates for the 2003 mathematics assessment are available at the NAEP web site (http://nces.ed.gov/nationsreportcard/mathematics/sampledesign.asp).

How Is Student Mathematics Performance Reported?

The results of student performance on the NAEP assessments are reported for various groups of students (e.g., fourth-grade female students or students who took the assessment in different years). NAEP does not produce scores for individual students, or report scores for schools. Nor are data produced for school districts, except that some large urban districts voluntarily participated in the assessment on a trial basis and were sampled as states were sampled. Mathematics performance for groups of students is reported in two ways: 1) average scale scores and 2) achievement levels.

Scale Scores: Student performance is reported as an average score based on the NAEP mathematics scale, which ranges from 0 to 500 and is linked to the corresponding scales in 1990, 1992, 1996, and 2000. Subscales were created to reflect performance on each of the five content areas defined in the NAEP mathematics framework. An overall composite scale was developed by weighting each of the mathematics subscales for the grade based on its relative importance in the framework. This composite scale is the metric used to present the average scale scores and selected percentiles used in NAEP reports.

Achievement Levels: Student mathematics performance is also reported in terms of three achievement levels—*Basic, Proficient*, and *Advanced*. Results based on achievement levels are expressed in terms of the percentage of students who attained each level. The three achievement levels are defined as follows:

- *Basic:* This level denotes partial mastery of prerequisite knowledge and skills that are fundamental for proficient work at each grade.
- *Proficient:* This level represents solid academic performance for each grade assessed. Students reaching this level have demonstrated competency over challenging subject matter, including subject-matter knowledge, application of such knowledge to real-world situations, and analytical skills appropriate to the subject matter.
- *Advanced:* This level signifies superior performance.

The achievement levels are performance standards adopted by the National Assessment Governing Board (NAGB) as part of its statutory responsibilities mandated by Congress. The levels represent collective judgments of what students should know and be able to do for each grade tested. They are based on recommendations made by broadly representative panels of classroom teachers, education specialists, and members of the general public. As provided by law, the National Center for Education Statistics (NCES), upon review of congressionally mandated evaluations of NAEP, has determined that the achievement levels are to be used on a trial basis until it is determined that the achievement levels are "reasonable, valid, and informative to the public."[1] However, both NCES and NAGB believe these performance standards are useful for understanding trends in student achievement. They have been widely used by national and state officials as a common yardstick for academic performance. The mathematics achievement-level descriptions are summarized in figure 1.

Students With Disabilities (SD) and/or Limited-English-Proficient (LEP) Students

The results displayed in this report and official publications of NAEP 2003 results are based on representative samples that include students with disabilities (SD) and limited-English-proficient students (LEP). Some of these students were assessed using accommodations that allowed them to participate. In state NAEP mathematics assessments prior to 2000, no testing accommodations or adaptations were permitted for special-needs students in these samples. However, research carried out by NAEP showed that the results for such accommodated students could be combined with the results for nonaccommodated students without compromising the validity of the NAEP scales in trend comparisons. Therefore, the special-needs students who typically received accommodations in their classroom testing, and who required these accommodations to participate, also received them in the NAEP assessment, provided the accommodations did not change the nature of what was tested.

In 2000, NAEP used a split sample of schools—one sample in which accommodations were permitted for special-needs students who normally received them and another sample in which accommodations were not permitted. Therefore, there are two different sets of results displayed for 2000. Results for the assessment years where accommodations were not permitted in state NAEP assessments (1990, 1992, 1996, and 2000) are reported in the same tables as the results where accommodations were permitted (2000 and 2003). The results labeled "Accommodations not permitted" are based on the same procedures as previously reported data. The results labeled "Accommodations permitted" for 2000 are based on the new procedures.

Statistical comparisons are made between the results across years, regardless of accommodation conditions, because NAEP's statistical studies showed that these comparisons could be made and the results remain valid. For 2000, when accommodations were permitted for one sample and not for another sample, comparisons to both samples are available in tables and in the data tool (http://nces.ed.gov/nationsreportcard/naepdata/). In the text of this report, comparisons to the 2000 results are discussed only for the sample for which accommodations were permitted.

Cautions in Interpreting Results

The averages and percentages in this report have a standard error—a range of up to a few points above or below the score—which takes into account potential score fluctuation due to sampling error and measurement error. Statistical tests that factor in these standard errors are used to determine whether the differences between average scores or percentages are significant. All differences were tested for statistical significance at the 0.05 level. NAEP sample sizes have increased since 2002 compared to previous years, resulting in smaller standard errors. As a consequence, smaller differences are detected as statistically significant than in previous assessments.

In this report, statistically significant differences are referred to as "significant differences" or "significantly different." Significant differences between 2003 and prior assessments are marked with a notation (*) in the tables. Any differences in scores within a year or across years that are mentioned in the text as "higher," "lower," "greater," or "smaller" are statistically significant.

Estimates based on small subgroups are likely to have large standard errors. Consequently some seemingly large differences may not be statistically significant. The reader is cautioned to rely on reported differences in the tables and/or text, which are statistically significant, rather than on the apparent magnitude of any difference. Readers are also cautioned against interpreting NAEP results causally. Inferences related to subgroup performance, for example, should take into account the many socioeconomic and educational factors that may affect student performance.

No Child Left Behind Act of 2001, Pub. L. No. 107–110, 115 Stat. 1425 (2001).

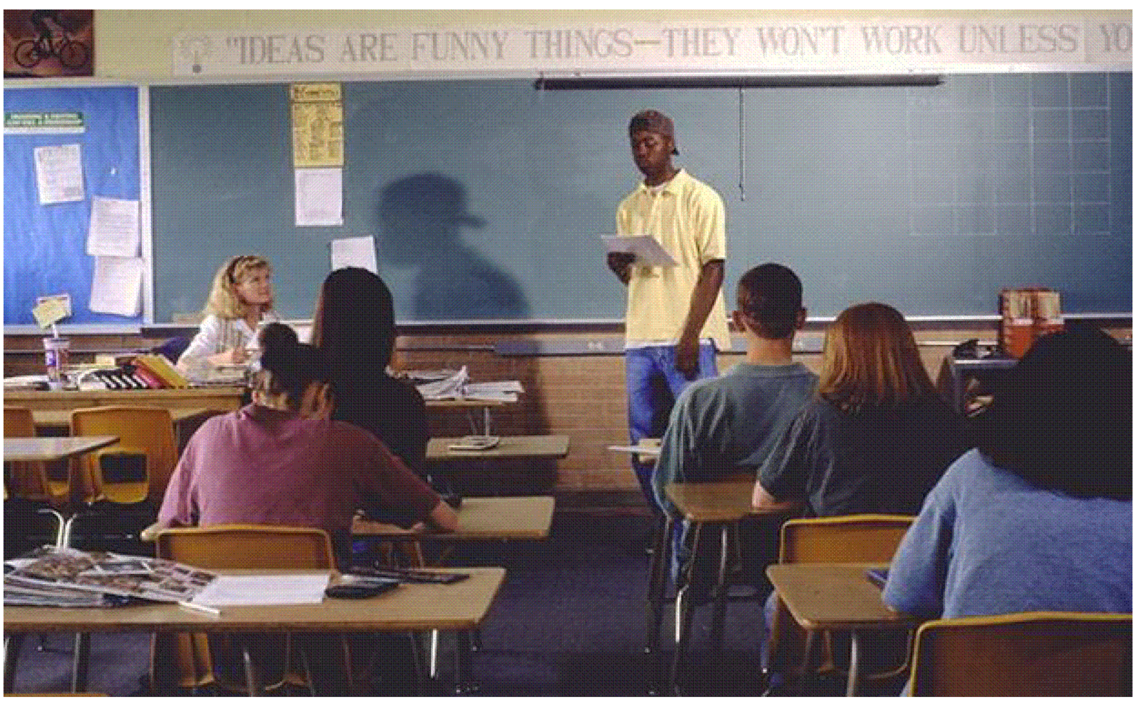


FIGURE 1

The Nation's Report Card 2003 State Assessment

Descriptions of NAEP mathematics achievement levels, grade 8

Basic Level (262)	Eighth-grade students performing at the *Basic* level should exhibit evidence of conceptual and procedural understanding in the five NAEP content areas. This level of performance signifies an understanding of arithmetic operations—including estimation—on whole numbers, decimals, fractions, and percents.

For example, eighth-graders performing at the *Basic* level should complete problems correctly with the help of structural prompts such as diagrams, charts, and graphs. They should be able to solve problems in all NAEP content areas through the appropriate selection and use of strategies and technological tools—including calculators, computers, and geometric shapes. Students at this level also should be able to use fundamental algebraic and informal geometric concepts in problem solving.

As they approach the *Proficient* level, students at the *Basic* level should be able to determine which of the available data are necessary and sufficient for correct solutions and use them in problem solving. However, these eighth-graders show limited skill in communicating mathematically.

Proficient Level (299)	Eighth-grade students performing at the *Proficient* level should apply mathematical concepts and procedures consistently to complex problems in the five NAEP content areas.

For example, eighth-graders performing at the *Proficient* level should be able to conjecture, defend their ideas, and give supporting examples. They should understand the connections among fractions, percents, decimals, and other mathematical topics such as algebra and functions. Students at this level are expected to have a thorough understanding of *Basic*-level arithmetic operations—an understanding sufficient for problem solving in practical situations.

Quantity and spatial relationships in problem solving and reasoning should be familiar to them, and they should be able to convey underlying reasoning skills beyond the level of arithmetic. They should be able to compare and contrast mathematical ideas and generate their own examples. These students should make inferences from data and graphs, apply properties of informal geometry, and accurately use the tools of technology. Students at this level should understand the process of gathering and organizing data and be able to calculate, evaluate, and communicate results within the domain of statistics and probability.

Advanced Level (333)	Eighth-grade students performing at the *Advanced* level should be able to reach beyond the recognition, identification, and application of mathematical rules in order to generalize and synthesize concepts and principles in the five NAEP content areas.

For example, eighth-graders performing at the *Advanced* level should be able to probe examples and counterexamples in order to shape generalizations from which they can develop models. Eighth-graders performing at the *Advanced* level should use number sense and geometric awareness to consider the reasonableness of an answer. They are expected to use abstract thinking to create unique problem-solving techniques and explain the reasoning processes underlying their conclusions.

NOTE: The scores in parentheses indicate the cutpoint on the scale at which the achievement-level range begins. SOURCE: National Assessment Governing Board. (2002). *Mathematics Framework for the 2003 National Assessment of Educational Progress.* Washington, DC: Author.

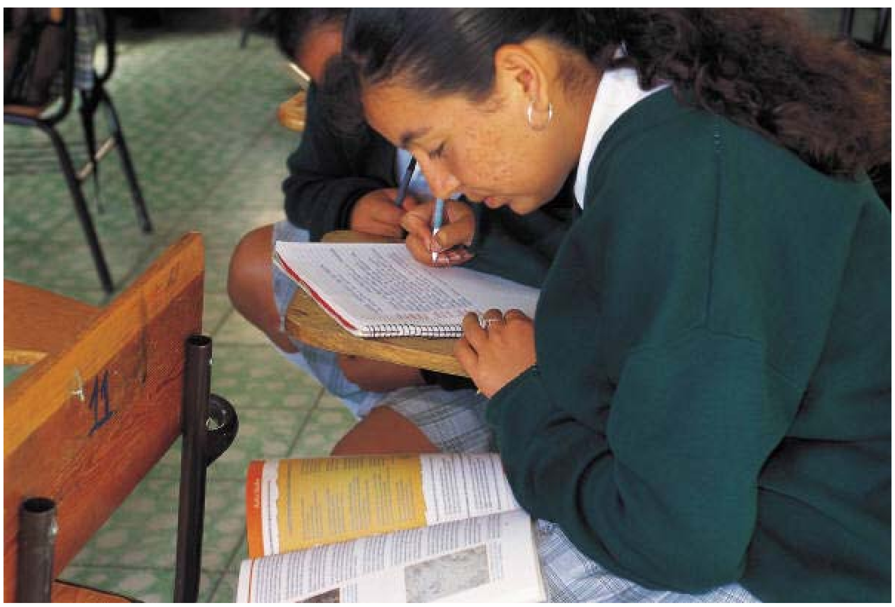
NAEP Mathematics 2003 Overall Scale Score and Achievement-Level Results for Public School Students

Overall Scale Score Results

In this section student performance is reported as an average score based on the NAEP mathematics scale, which ranges from 0 to 500. Scores on this scale are comparable from 1990 through 2003.

Prior to 2000, testing accommodations were not provided for students with special needs in state mathematics assessments.

Table 1 shows the overall performance results of grade 8 public school students in New Hampshire and the nation. The first column of results presents the average score on the NAEP mathematics scale. The subsequent columns show the score at selected percentiles. The percentile indicates the percentage of students who performed below the score for that percentile. For example, 10 percent of the students had scores that were lower than the score shown for the 10th percentile.



Grade 8 Scale Score Results

- In 2003, the average scale score for students in New Hampshire was 286. This was higher than that of students across the nation (276).

- In New Hampshire, the average scale score for students in 2003 was higher than that in 1990 (273).

- In New Hampshire, the average scale score for students in 2003 was higher than that in 1992 (278). Similarly, the average scale score for students across the nation in 2003 was higher than that in 1992 (267).

TABLE 1

The Nation's Report Card 2003 State Assessment

Average mathematics scale scores and selected percentiles, grade 8 public schools: 1990–2003

	Average Scale Score	Scale score distribution				
		10th Percentile	25th Percentile	50th Percentile	75th Percentile	90th Percentile
Accommodations not permitted						
1990 New Hampshire	273 (0.9)*	234 (1.1)*	253 (1.8)*	273 (1.0)*	294 (1.3)*	313 (2.3)*
Nation (Public)	262 (1.4)*	214 (1.8)*	237 (1.4)*	263 (1.5)*	288 (1.7)*	307 (1.8)*
1992 New Hampshire	278 (1.0)*	239 (1.2)*	259 (0.8)*	279 (0.8)*	299 (1.4)*	316 (1.7)*
Nation (Public)	267 (1.0)*	219 (1.5)*	242 (1.5)*	268 (1.1)*	293 (1.3)*	314 (1.6)*
Accommodations permitted						
2003 New Hampshire	286 (0.8)	246 (1.1)	266 (1.4)	287 (1.1)	308 (1.5)	326 (1.6)
Nation (Public)	276 (0.3)	228 (0.6)	253 (0.4)	278 (0.4)	301 (0.3)	321 (0.3)

* Value is significantly different from the value for the same jurisdiction in 2003.
NOTE: The NAEP mathematics scale ranges from 0 to 500. The standard errors of the statistics in the table appear in parentheses. All differences were tested for statistical significance at the 0.05 level using unrounded numbers. Performance comparisons may be affected by differences in exclusion rates for students with disabilities and limited-English-proficient students in the NAEP samples and changes in sample sizes. NAEP sample sizes have increased since 2002 compared to previous years, resulting in smaller detectable differences than in previous assessments. In addition to allowing for accommodations, the accommodations-permitted results for national public schools (2000 and 2003) differ slightly from previous years' results, and from previously reported results for 2000, due to changes in sample weighting procedures.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1990–2003 Mathematics Assessments.

Overall Achievement-Level Results

In this section student performance is reported as the percentage of students performing relative to standards set by the National Assessment Governing Board (NAGB). These performance standards for what students should know and be able to do were based on the recommendations of broadly representative panels of educators and members of the public.

Table 2 presents the percentage of students at grade 8 who performed below *Basic*, at or above *Basic*, at or above *Proficient*, and at the *Advanced* level. Because the percentages are cumulative from *Basic* to *Proficient* to *Advanced*, they sum to more than 100 percent. Only the percentage of students performing at or above *Basic* (which includes the students at *Proficient* and *Advanced*) plus the students below *Basic* will sum to 100 percent (except for rounding).

Grade 8 Achievement-Level Results

- In 2003, the percentage of New Hampshire's students who performed at or above the *Proficient* level was 35 percent. This was greater than the percentage of the nation's public school students who performed at or above *Proficient* (27 percent).

- In New Hampshire, the percentage of students who performed at or above the *Proficient* level in 2003 was greater than that in 1990 (20 percent).

- In New Hampshire, the percentage of students who performed at or above the *Proficient* level in 2003 was greater than that in 1992 (25 percent).

TABLE 2

The Nation's Report Card 2003 State Assessment

Percentage of students at or above each mathematics achievement level, grade 8 public schools: 1990–2003

	Below *Basic*	At or above *Basic*	At or above *Proficient*	At or above *Advanced*
Accommodations not permitted				
1990 New Hampshire	35 (1.5)*	65 (1.5)*	20 (1.2)*	3 (0.5)*
Nation (Public)	49 (1.5)*	51 (1.5)*	15 (1.1)*	2 (0.4)*
1992 New Hampshire	29 (1.3)*	71 (1.3)*	25 (1.4)*	3 (0.5)*
Nation (Public)	44 (1.2)*	56 (1.2)*	20 (1.0)*	3 (0.4)*
Accommodations permitted				
2003 New Hampshire	21 (1.1)	79 (1.1)	35 (1.2)	7 (0.8)
Nation (Public)	33 (0.3)	67 (0.3)	27 (0.3)	5 (0.1)

* Value is significantly different from the value for the same jurisdiction in 2003.

NOTE: The standard errors of the statistics in the table appear in parentheses. Achievement levels correspond to the following points on the NAEP mathematics scale: below *Basic*, 261 or lower; *Basic*, 262-298; *Proficient*, 299-332; and *Advanced*, 333 and above. All differences were tested for statistical significance at the 0.05 level using unrounded numbers. Details may not sum to totals due to rounding. Performance comparisons may be affected by differences in exclusion rates for students with disabilities and limited-English-proficient students in the NAEP samples and changes in sample sizes. NAEP sample sizes have increased since 2002 compared to previous years, resulting in smaller detectable differences than in previous assessments. In addition to allowing for accommodations, the accommodations-permitted results for national public schools (2000 and 2003) differ slightly from previous years' results, and from previously reported results for 2000, due to changes in sample weighting procedures.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1990–2003 Mathematics Assessments.

Comparisons Between New Hampshire and Other Participating States and Jurisdictions

In 2003, 53 jurisdictions participated in the mathematics assessment. These include the 50 states, the District of Columbia and the two groups of Department of Defense Education Activity (DoDEA) schools: Domestic Dependent Elementary and Secondary Schools (DDESS) and Department of Defense Dependents Schools (DoDDS).

Comparisons by Average Scale Scores

Figure 2 compares New Hampshire's 2003 overall mathematics scale scores at grade 8 with those of all other participating states and jurisdictions. The different shadings indicate whether a state's or jurisdiction's average scale score was found to be higher than, lower than, or not significantly different from that of New Hampshire in the NAEP 2003 mathematics assessment.

Grade 8 Scale Score Comparisons Results

- Students' scale scores in New Hampshire were higher than those in 40 jurisdictions, not significantly different from those in 11 jurisdictions, and lower than those in 1 jurisdiction.

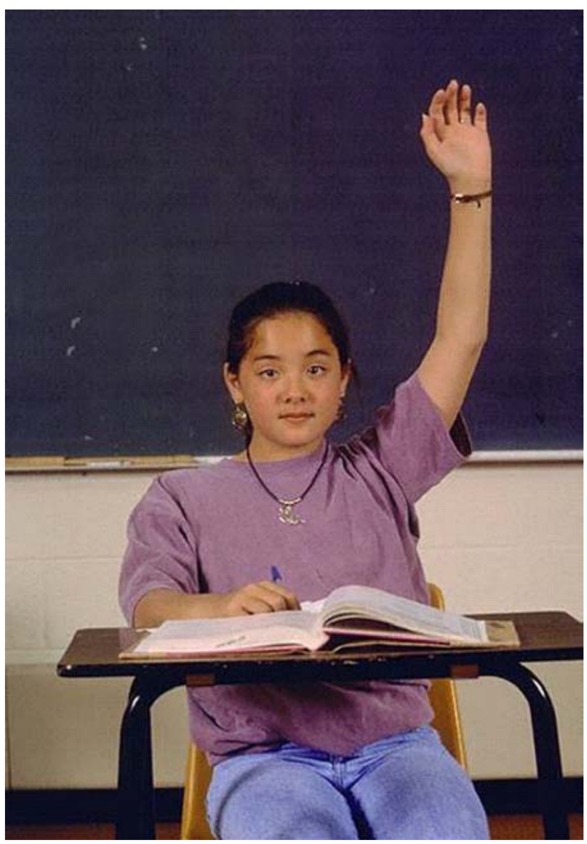
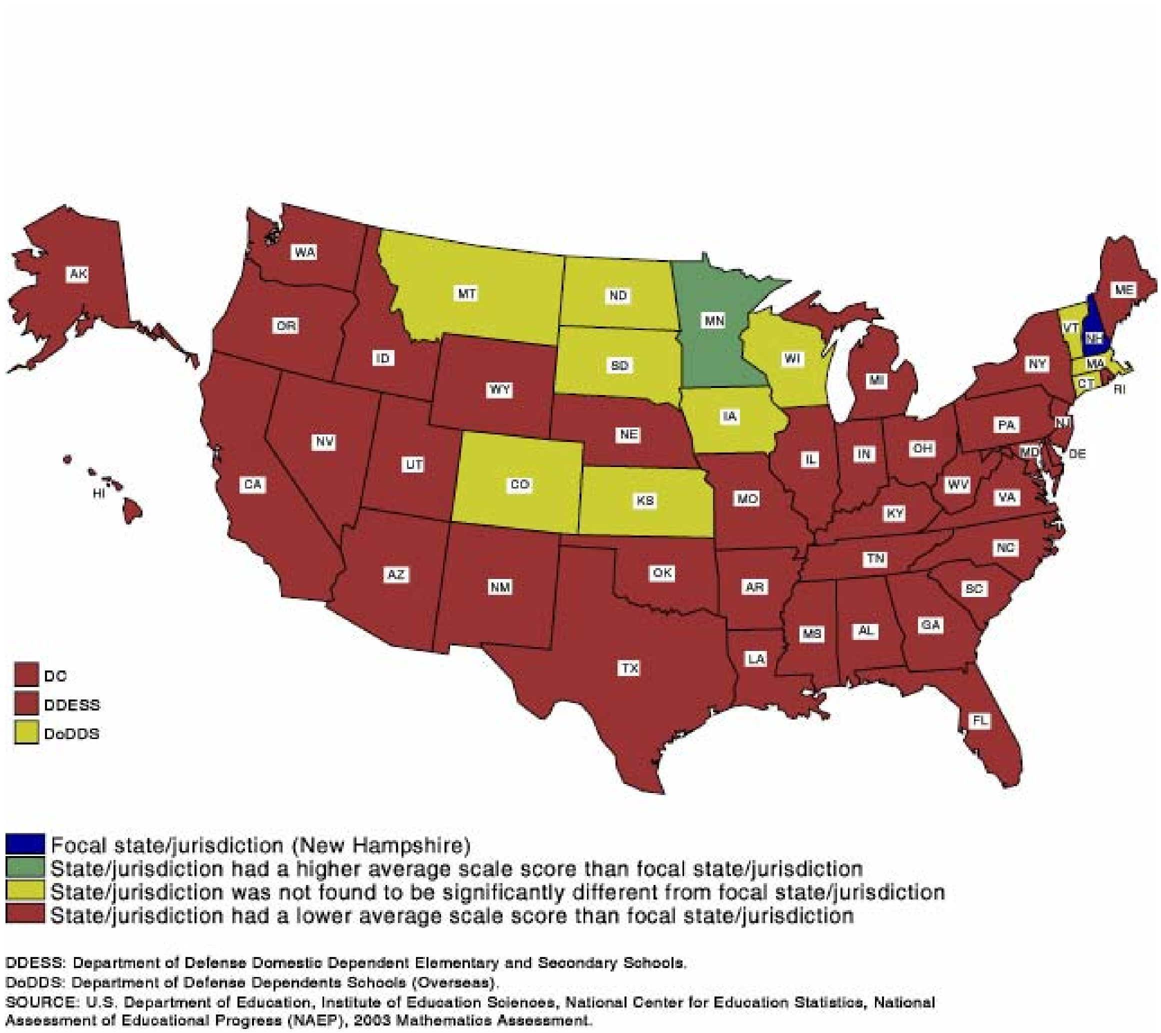


FIGURE 2

The Nation's Report Card 2003 State Assessment

New Hampshire's average mathematics scale score compared with scores for other participating jurisdictions, grade 8 public schools: 2003



Comparisons by Achievement Levels

Figure 3 permits comparisons of all jurisdictions participating in the NAEP 2003 mathematics assessment in terms of percentages of grade 8 students performing at or above the *Proficient* level. The participating states and jurisdictions are grouped into categories reflecting student performance compared to that in New Hampshire. The jurisdictions are grouped by whether the percentage of their students with scores at or above the *Proficient* level (including *Advanced*) was found to be higher than, not significantly different from, or lower than the percentage in New Hampshire. Note that the arrangement of the states and the other jurisdictions within each category is alphabetical; statistical comparisons among jurisdictions in each of the three categories are not included in this report.

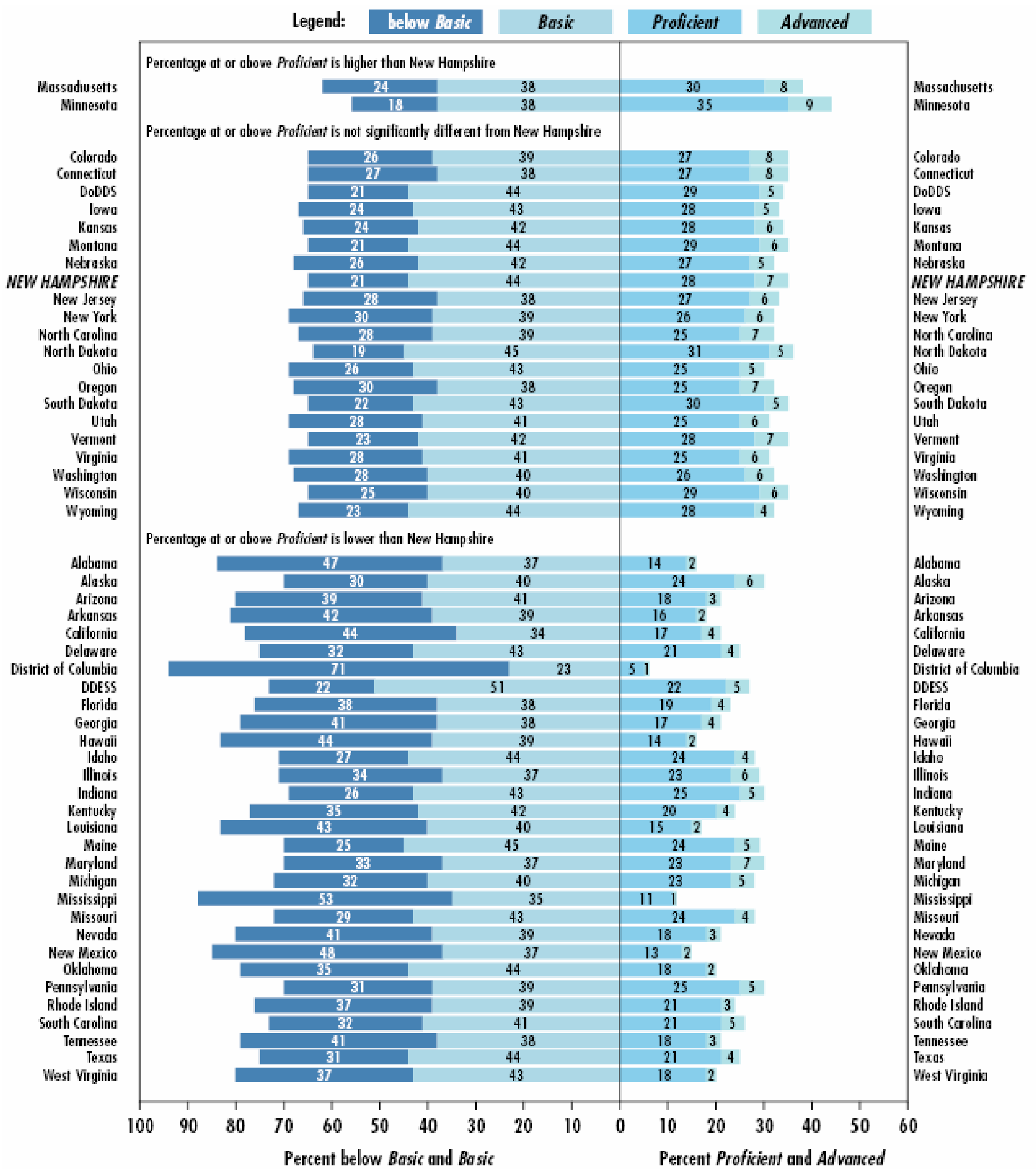
Grade 8 Achievement-Level Comparisons Results

- At grade 8, 2 jurisdictions had higher percentages of students at or above the *Proficient* level than that of New Hampshire, 20 jurisdictions had percentages that were not significantly different from that of New Hampshire, and 30 jurisdictions had lower percentages than that of New Hampshire.

FIGURE 3

The Nation's Report Card 2003 State Assessment

Percentage of students within each mathematics achievement-level range, and New Hampshire's percentage at or above Proficient compared with other participating jurisdictions, grade 8 public schools: By state, 2003



DDESS: Department of Defense Domestic Dependent Elementary and Secondary Schools.
DoDDS: Department of Defense Dependents Schools (Overseas).
NOTE: The bars above contain percentages of students in each NAEP mathematics achievement range. Achievement levels corresponding to each population of students are aligned at the point where the *Proficient* category begins, so that they may be compared at *Proficient* and above. Detail may not sum to totals because of rounding.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Mathematics Assessment.

Mathematics Performance by Demographic Characteristics

This section of the report presents trend results for students in New Hampshire and the nation by demographic characteristics. Student performance data are reported for:

- Gender
- Race/ethnicity
- Eligibility for free/reduced-price school lunch
- Type of location (2000 and later)

Definitions of NAEP reporting groups are available on the NAEP web site (http://nces.ed.gov/nationsreportcard/mathematics/results2003/interpret-results.asp#RepGroups).

Each of the variables is reported in tables that present the percentage of students belonging to each subgroup in the first column and the average scale score in the second column. The columns to the right show the percentage of students at or above each achievement-level.

The reader is cautioned against making causal inferences about the performance of groups of students relative to demographic variables. Many factors other than those discussed here, including home and school factors, may affect student performance.

NAEP collects information on many additional variables, including school and home factors related to achievement. All of this information is in an interactive database available on the NAEP web site (http://nces.ed.gov/nationsreportcard/naepdata/).

Gender

Information on student gender is reported by schools on rosters of students eligible to be assessed.

Table 3 shows scale scores and achievement-level data for public-school students at grade 8 in New Hampshire and the nation by gender.

Grade 8 Scale Score Results by Gender

- In New Hampshire, male students' average scale score was 287 in 2003. This was not found to differ significantly from that of female students (286).
- In 2003, male students in New Hampshire had an average scale score in Mathematics (287) that was higher than that of male students across the nation (277). Female students in New Hampshire had an average score (286) that was higher than that of female students nationwide (275).
- In New Hampshire, the average scale scores of both males and females were higher in 2003 than in 1990.
- In New Hampshire, the average scale scores of both males and females were higher in 2003 than in 1992.

Grade 8 Achievement-Level Results by Gender

- In 2003, 36 percent of males and 33 percent of females performed at or above the *Proficient* level in New Hampshire. The difference between these percentages was not significant.
- The percentage of males in New Hampshire's public schools who were at or above the *Proficient* level in 2003 (36 percent) was greater than that of males in the nation (29 percent).
- The percentage of females in New Hampshire's public schools who were at or above the *Proficient* level in 2003 (33 percent) was greater than that of females in the nation (26 percent).
- In New Hampshire, the percentages of both males and females performing at or above the *Proficient* level were greater in 2003 than in 1990.
- In New Hampshire, the percentages of both males and females performing at or above the *Proficient* level were greater in 2003 than in 1992.

TABLE 3

The Nation's Report Card 2003 State Assessment

Average mathematics scale scores and percentage of students at or above each achievement level, by gender, grade 8 public schools: 1990–2003

	Percentage of Students	Average Scale Score	Below *Basic*	At or above *Basic*	At or above *Proficient*	At *Advanced*
Male						
Accommodations not permitted						
1990 New Hampshire	53 (1.1)	273 (1.1)*	36 (2.2)*	64 (2.2)*	20 (1.6)*	3 (0.7)*
Nation (Public)	51 (1.1)	262 (1.7)*	49 (2.0)*	51 (2.0)*	17 (1.5)*	2 (0.5)*
1992 New Hampshire	50 (1.0)	279 (1.3)*	28 (1.6)*	72 (1.6)*	26 (1.9)*	3 (0.7)*
Nation (Public)	52 (0.6)	266 (1.1)*	45 (1.5)*	55 (1.5)*	20 (1.3)*	3 (0.5)*
Accommodations permitted						
2003 New Hampshire	51 (1.1)	287 (1.0)	21 (1.1)	79 (1.1)	36 (1.4)	7 (1.0)
Nation (Public)	50 (0.2)	277 (0.3)	33 (0.4)	67 (0.4)	29 (0.3)	6 (0.2)
Female						
Accommodations not permitted						
1990 New Hampshire	47 (1.1)	274 (1.2)*	35 (1.9)*	65 (1.9)*	21 (1.8)*	3 (0.7)*
Nation (Public)	49 (1.1)	261 (1.4)*	49 (1.7)*	51 (1.7)*	14 (1.2)*	2 (0.5)*
1992 New Hampshire	50 (1.0)	278 (1.2)*	29 (1.6)*	71 (1.6)*	24 (1.5)*	3 (0.6)*
Nation (Public)	48 (0.6)	267 (1.1)*	44 (1.5)*	56 (1.5)*	20 (1.3)*	3 (0.5)*
Accommodations permitted						
2003 New Hampshire	49 (1.1)	286 (1.1)	22 (1.7)	78 (1.7)	33 (1.7)	6 (0.9)
Nation (Public)	50 (0.2)	275 (0.3)	34 (0.4)	66 (0.4)	26 (0.3)	4 (0.1)

* Value is significantly different from the value for the same jurisdiction in 2003.
NOTE: The NAEP mathematics scale ranges from 0 to 500. The standard errors of the statistics in the table appear in parentheses. Achievement levels correspond to the following points on the NAEP mathematics scale: below *Basic*, 261 or lower; *Basic*, 262-298; *Proficient*, 299-332; and *Advanced*, 333 and above. All differences were tested for statistical significance at the 0.05 level using unrounded numbers. Details may not sum to totals due to rounding. Performance comparisons may be affected by differences in exclusion rates for students with disabilities and limited-English-proficient students in the NAEP samples and changes in sample sizes. NAEP sample sizes have increased since 2002 compared to previous years, resulting in smaller detectable differences than in previous assessments. In addition to allowing for accommodations, the accommodations-permitted results for national public schools (2000 and 2003) differ slightly from previous years' results, and from previously reported results for 2000, due to changes in sample weighting procedures.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1990–2003 Mathematics Assessments.

Race/Ethnicity

Schools report the racial/ethnic subgroup that best described the students eligible to be assessed. The five mutually exclusive categories are White, Black, Hispanic, Asian/Pacific Islander, and American Indian/Alaska Native.

Table 4 shows scale scores and achievement-level data for public-school students at grade 8 in New Hampshire and the nation by race/ethnicity.

Grade 8 Scale Score Results by Race/Ethnicity

- The average scale score of White students in New Hampshire was higher in 2003 than in 1990.
- The average scale score of White students in New Hampshire was higher in 2003 than in 1992.

Grade 8 Achievement-Level Results by Race/Ethnicity

- The percentage of White students in New Hampshire performing at or above the *Proficient* level was greater in 2003 than in 1990.
- The percentage of White students in New Hampshire performing at or above the *Proficient* level was greater in 2003 than in 1992.

TABLE 4

Average mathematics scale scores and percentage of students at or above each achievement level, by race/ethnicity, grade 8 public schools: 1990–2003

	Percentage of Students	Average Scale Score	Below *Basic*	At or above *Basic*	At or above *Proficient*	At *Advanced*
White						
Accommodations not permitted						
1990 New Hampshire	98 (0.5)*	273 (0.9)*	35 (1.5)*	65 (1.5)*	20 (1.1)*	3 (0.5)*
Nation (Public)	73 (0.8)*	269 (1.4)*	41 (1.7)*	59 (1.7)*	18 (1.4)*	3 (0.5)*
1992 New Hampshire	96 (1.7)	278 (0.9)*	29 (1.3)*	71 (1.3)*	25 (1.3)*	3 (0.5)*
Nation (Public)	72 (0.6)*	276 (1.1)*	34 (1.4)*	66 (1.4)*	25 (1.2)*	3 (0.5)*
Accommodations permitted						
2003 New Hampshire	95 (0.5)	287 (0.8)	20 (1.2)	80 (1.2)	35 (1.2)	7 (0.8)
Nation (Public)	62 (0.4)	287 (0.3)	21 (0.3)	79 (0.3)	36 (0.4)	7 (0.2)
Black						
Accommodations not permitted						
1990 New Hampshire	# (0.1)!*	--- (---)	--- (---)	--- (---)	--- (---)	--- (---)
Nation (Public)	16 (0.5)	236 (2.8)*	79 (2.4)*	21 (2.4)*	5 (1.1)	# (***)
1992 New Hampshire	1 (0.2)!	--- (---)	--- (---)	--- (---)	--- (---)	--- (---)
Nation (Public)	17 (0.3)	236 (1.3)*	81 (2.0)*	19 (2.0)*	2 (0.7)*	# (***)
Accommodations permitted						
2003 New Hampshire	1 (0.2)	--- (---)	--- (---)	--- (---)	--- (---)	--- (---)
Nation (Public)	17 (0.3)	252 (0.5)	61 (0.9)	39 (0.9)	7 (0.3)	# (0.1)
Hispanic						
Accommodations not permitted						
1990 New Hampshire	1 (0.2)!*	--- (---)	--- (---)	--- (---)	--- (---)	--- (---)
Nation (Public)	7 (0.5)*	245 (4.4)*	67 (4.5)*	33 (4.5)*	7 (2.1)	1 (0.4)
1992 New Hampshire	1 (0.2)!*	--- (---)	--- (---)	--- (---)	--- (---)	--- (---)
Nation (Public)	8 (0.4)*	247 (1.2)*	67 (2.0)*	33 (2.0)*	6 (1.0)*	# (0.2)*
Accommodations permitted						
2003 New Hampshire	2 (0.2)	--- (---)	--- (---)	--- (---)	--- (---)	--- (---)
Nation (Public)	15 (0.3)	258 (0.6)	53 (0.9)	47 (0.9)	11 (0.5)	1 (0.1)
Asian/Pacific Islander						
Accommodations not permitted						
1990 New Hampshire	1 (0.2)!	--- (---)	--- (---)	--- (---)	--- (---)	--- (---)
Nation (Public)	2 (0.5)!*	275 (5.7)!*	36 (5.4)!*	64 (5.4)!*	30 (6.8)!	6 (3.1)!
1992 New Hampshire	1 (0.2)!	--- (---)	--- (---)	--- (---)	--- (---)	--- (---)
Nation (Public)	2 (0.3)*	290 (7.0)	25 (5.8)	75 (5.8)	43 (8.0)	14 (4.9)
Accommodations permitted						
2003 New Hampshire	1 (0.3)	--- (---)	--- (---)	--- (---)	--- (---)	--- (---)
Nation (Public)	4 (0.2)	289 (1.3)	23 (1.2)	77 (1.2)	42 (1.4)	12 (1.4)

Footnotes appear at the bottom of the last page of this table.

TABLE 4

The Nation's Report Card 2003 State Assessment

Average mathematics scale scores and percentage of students at or above each achievement level, by race/ethnicity, grade 8 public schools: 1990–2003 (continued)

	Percentage of Students	Average Scale Score	Below *Basic*	At or above *Basic*	At or above *Proficient*	At *Advanced*
American Indian						
Accommodations not permitted						
1990 New Hampshire	# (0.2)!	--- (---)	--- (---)	--- (---)	--- (---)	--- (---)
Nation (Public)	1 (0.7)!	--- (---)	--- (---)	--- (---)	--- (---)	--- (---)
1992 New Hampshire	# (***)	--- (---)	--- (---)	--- (---)	--- (---)	--- (---)
Nation (Public)	1 (0.2)!	--- (---)	--- (---)	--- (---)	--- (---)	--- (---)
Accommodations permitted						
2003 New Hampshire	# (0.1)!	--- (---)	--- (---)	--- (---)	--- (---)	--- (---)
Nation (Public)	1 (0.1)	265 (1.2)	46 (1.8)	54 (1.8)	16 (1.3)	2 (0.7)

--- Reporting standards are not met. Sample size is insufficient to permit a reliable estimate.
Estimate rounds to zero.
* Value is significantly different from the value for the same jurisdiction in 2003.
(***) Standard error estimate cannot be accurately determined.
! The nature of the sample does not allow accurate determination of the variability of the statistic.
NOTE: The NAEP mathematics scale ranges from 0 to 500. The standard errors of the statistics in the table appear in parentheses. Achievement levels correspond to the following points on the NAEP mathematics scale: below *Basic*, 261 or lower; *Basic*, 262-298; *Proficient*, 299-332; and *Advanced*, 333 and above. All differences were tested for statistical significance at the 0.05 level using unrounded numbers. Details may not sum to totals due to rounding. Performance comparisons may be affected by differences in exclusion rates for students with disabilities and limited-English-proficient students in the NAEP samples and changes in sample sizes. NAEP sample sizes have increased since 2002 compared to previous years, resulting in smaller detectable differences than in previous assessments. In addition to allowing for accommodations, the accommodations-permitted results for national public schools (2000 and 2003) differ slightly from previous years' results, and from previously reported results for 2000, due to changes in sample weighting procedures.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1990–2003 Mathematics Assessments.

Free/Reduced-Price Lunch Eligibility

NAEP collects data on eligibility for the federal program providing free or reduced-price school lunches. The free/reduced-price lunch component of the National School Lunch Program (NSLP) offered through the U.S. Department of Agriculture (USDA) is designed to ensure that children near or below the poverty line receive nourishing meals. This program is available to public schools, nonprofit private schools, and residential child-care institutions. Eligibility is determined through the USDA's Income Eligibility Guidelines, and results for this category of students are included as an indicator of poverty. NAEP first collected information on participation in this program in 1996.

Table 5 shows scale scores and achievement-level data for public-school students at grade 8 in New Hampshire and the nation by eligibility for free/reduced-price lunch.

Grade 8 Scale Score Results by Free/Reduced-Price Lunch Eligibility

- Students in New Hampshire eligible for free/reduced-price lunch had an average Mathematics scale score of 268. This was lower than that of students in New Hampshire not eligible for this program (289).

- Students in New Hampshire eligible for free/reduced-price lunch had an average scale score (268) that was higher than that of students in the nation who were eligible (258).

Grade 8 Achievement-Level Results by Free/Reduced-Price Lunch Eligibility

- In New Hampshire, 16 percent of students who were eligible for free/reduced-price lunch and 38 percent of those who were not eligible for this program performed at or above the *Proficient* level. These percentages were found to be significantly different from one another.

- For students in New Hampshire who were eligible for free/reduced-price lunch, the percentage at or above the *Proficient* level (16 percent) was not found to be significantly different from the corresponding percentage for their counterparts around the nation (11 percent).

TABLE 5

The Nation's Report Card 2003 State Assessment

Average mathematics scale scores and percentage of students at or above each achievement level, by eligibility for free/reduced-price school lunch, grade 8 public schools: 2003

	Percentage of Students	Average Scale Score	Below *Basic*	At or above *Basic*	At or above *Proficient*	At *Advanced*
Eligible						
Accommodations permitted						
2003 New Hampshire	13 (0.9)	268 (2.1)	42 (3.3)	58 (3.3)	16 (2.5)	2 (0.9)
Nation (Public)	36 (0.4)	258 (0.3)	53 (0.5)	47 (0.5)	11 (0.3)	1 (0.1)
Not Eligible						
Accommodations permitted						
2003 New Hampshire	79 (1.1)	289 (0.9)	18 (1.2)	82 (1.2)	38 (1.4)	7 (0.9)
Nation (Public)	58 (0.6)	287 (0.3)	22 (0.3)	78 (0.3)	37 (0.4)	7 (0.2)
Information Not Available						
Accommodations permitted						
2003 New Hampshire	8 (0.9)	286 (2.4)	22 (3.0)	78 (3.0)	36 (3.1)	6 (2.0)
Nation (Public)	6 (0.4)	278 (1.3)	32 (1.3)	68 (1.3)	29 (1.5)	6 (0.6)

NOTE: The NAEP mathematics scale ranges from 0 to 500. The standard errors of the statistics in the table appear in parentheses. Achievement levels correspond to the following points on the NAEP mathematics scale: below *Basic*, 261 or lower; *Basic*, 262-298; *Proficient*, 299-332; and *Advanced*, 333 and above. All differences were tested for statistical significance at the 0.05 level using unrounded numbers. Details may not sum to totals due to rounding. Performance comparisons may be affected by differences in exclusion rates for students with disabilities and limited-English-proficient students in the NAEP samples and changes in sample sizes. NAEP sample sizes have increased since 2002 compared to previous years, resulting in smaller detectable differences than in previous assessments. In addition to allowing for accommodations, the accommodations-permitted results for national public schools (2000 and 2003) differ slightly from previous years' results, and from previously reported results for 2000, due to changes in sample weighting procedures.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Mathematics Assessment.

Type of Location

Schools that participated in the assessment were classified into three mutually exclusive types of community in which the school is located: central city, urban fringe/large town, and rural/small town. These categories indicate the geographic locations of schools. Central city is geographical term meaning the largest city of a Metropolitan Statistical Area and is not synonymous with "inner city."

Recently, the National Center for Education Statistics (NCES) introduced new methods to identify the type of location assigned to each school in the Common Core of Data (CCD). The new methods were put into place by NCES in order to improve the quality of the assignments, and they take into account more information about the exact physical location of the school. The variable was revised in NAEP beginning with the 2000 assessment; therefore, results are not presented for assessment years prior to 2000.

Table 6 shows scale scores and achievement-level data for public-school students at grade 8 in New Hampshire and the nation by type of location.

Grade 8 Scale Score Results by Type of Location

- In 2003, in New Hampshire, the average scale score of students attending schools in central cities was not found to differ significantly from that of students in urban fringes/large towns or rural areas/small towns.

Grade 8 Achievement-Level Results by Type of Location

- In 2003, the percentage of students attending schools in central cities in New Hampshire who performed at or above the *Proficient* level was not found to differ significantly from the corresponding percentages for students in urban fringes/large towns and rural areas/small towns.

TABLE 6

The Nation's Report Card 2003 State Assessment

Average mathematics scale scores and percentage of students at or above each achievement level, by type of location, grade 8 public schools: 2003

	Percentage of Students	Average Scale Score	Below *Basic*	At or above *Basic*	At or above *Proficient*	At *Advanced*
Central City						
Accommodations permitted						
2003 New Hampshire	17 (0.8)	283 (1.6)	25 (2.2)	75 (2.2)	31 (2.6)	6 (1.2)
Nation (Public)	27 (0.3)	267 (0.5)	44 (0.7)	56 (0.7)	20 (0.5)	4 (0.2)
Urban Fringe/Large Town						
Accommodations permitted						
2003 New Hampshire	24 (0.7)	287 (1.1)	21 (2.0)	79 (2.0)	37 (2.6)	7 (1.2)
Nation (Public)	42 (0.4)	280 (0.5)	29 (0.5)	71 (0.5)	31 (0.5)	6 (0.3)
Rural/Small Town						
Accommodations permitted						
2003 New Hampshire	59 (1.1)	287 (1.2)	20 (1.3)	80 (1.3)	35 (1.7)	7 (1.1)
Nation (Public)	31 (0.4)	279 (0.4)	29 (0.5)	71 (0.5)	28 (0.4)	4 (0.2)

NOTE: The NAEP mathematics scale ranges from 0 to 500. The standard errors of the statistics in the table appear in parentheses. Achievement levels correspond to the following points on the NAEP mathematics scale: below *Basic*, 261 or lower; *Basic*, 262-298; *Proficient*, 299-332; and *Advanced*, 333 and above. All differences were tested for statistical significance at the 0.05 level using unrounded numbers. Details may not sum to totals due to rounding. Performance comparisons may be affected by differences in exclusion rates for students with disabilities and limited-English-proficient students in the NAEP samples and changes in sample sizes. NAEP sample sizes have increased since 2002 compared to previous years, resulting in smaller detectable differences than in previous assessments. In addition to allowing for accommodations, the accommodations-permitted results for national public schools (2000 and 2003) differ slightly from previous years' results, and from previously reported results for 2000, due to changes in sample weighting procedures.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Mathematics Assessment.

Toward a More Inclusive NAEP

NAEP endeavors to assess all students selected in the randomized sampling process, including students with disabilities (SD) as well as students who are classified by their schools as limited-English-proficient (LEP). Some students sampled for participation in NAEP can be excluded from the sample according to carefully defined criteria. School personnel, guided by the student's Individualized Education Program (IEP), as well as eligibility for Section 504 services, make decisions regarding inclusion in the assessment of students with disabilities. They also make decisions regarding inclusion of LEP students, based on NAEP's guidelines. This includes evaluating the student's capability of participating in the assessment in English, as well as taking into consideration the number of years the student has been receiving instruction in English.

Percentages of students excluded from NAEP may vary considerably across states, and within a state, across years. Comparisons of results across states and within a state across years should be interpreted with caution if the exclusion rates vary widely. The percentages of students classified as SD or LEP in all participating states and jurisdictions are available in an interactive database at the NAEP web site (http://nces.ed.gov/nationsreportcard/naepdata/).

The results displayed in this report and in other publications of the NAEP 2003 mathematics results are based on representative samples that include SD and LEP students who were assessed either with or without accommodations, based on NAEP's guidelines. Prior to 2000, however, in state NAEP mathematics assessments no testing accommodations or adaptations were made available to the special-needs students in the samples that served as the basis for reported results.

In the 1996 national and 2000 national and state mathematics assessments, NAEP drew a second representative sample of schools. Accommodations were made available for students in this sample who required them, provided the accommodation did not change the nature of what was tested. For example, students could be assessed one-on-one or in small groups, receive extended time, or use a large-print test book. In mathematics, students had the option of using a bilingual English–Spanish test book. However, for mathematics students were not allowed to use calculators for any questions on which calculators were not permitted. NAEP has used these comparable samples to study the effects of allowing accommodations for special-needs students in the assessments. A series of technical research papers covering various NAEP subject areas has been published with the results of these comparisons (see http://nces.ed.gov/nationsreportcard/about/inclusion.asp #research).

Table 7 displays the percentages of special-needs students identified, excluded, and accommodated at grade 8.

Table 8 presents the total number of students assessed, the percentage of students sampled that were excluded, and average scale scores for all participating states and other jurisdictions at grade 8.

TABLE 7

The Nation's Report Card 2003 State Assessment

Percentage of SD and LEP students in mathematics assessments identified, excluded, and assessed, grade 8 public schools: 1990–2003

	SD and/or LEP		SD		LEP	
	New Hampshire	Nation (Public)	New Hampshire	Nation (Public)	New Hampshire	Nation (Public)
Accommodations not permitted						
1990 Identified	12 (0.8)	— (—)	12 (0.8)	— (—)	# (0.1)	— (—)
Excluded	4 (0.4)	— (—)	4 (0.4)	— (—)	# (0.1)	— (—)
Assessed under standard conditions	8 (0.7)	— (—)	7 (0.7)	— (—)	# (0.1)	— (—)
1992 Identified	12 (0.7)	10 (0.5)	12 (0.7)	8 (0.5)	# (0.1)	2 (0.2)
Excluded	5 (0.4)	6 (0.4)	5 (0.4)	5 (0.3)	# (0.1)	2 (0.2)
Assessed under standard conditions	7 (0.7)	4 (0.5)	7 (0.7)	3 (0.5)	# (0.1)	1 (0.1)
Accommodations permitted						
2003 Identified	20 (0.7)	19 (0.2)	19 (0.7)	14 (0.2)	1 (0.2)	6 (0.2)
Excluded	3 (0.5)	4 (0.1)	3 (0.5)	3 (0.1)	# (0.1)	1 (0.1)
Assessed under standard conditions	6 (0.5)	8 (0.2)	6 (0.5)	5 (0.1)	# (0.1)	4 (0.2)
Assessed with accommodations	10 (0.6)	7 (0.1)	9 (0.6)	6 (0.1)	1 (0.2)	1 (0.1)

— Not available.
Estimate rounds to zero.
(***) Standard error estimate cannot be accurately determined.
SD: Students with Disabilities. LEP: Limited-English-proficient students.
NOTE: The standard errors of the statistics in the table appear in parentheses. Detail may not sum to totals because of rounding. Some students were identified as both SD and LEP. Such students would be included in both the SD and LEP portions of the table.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1990–2003 Mathematics Assessments.

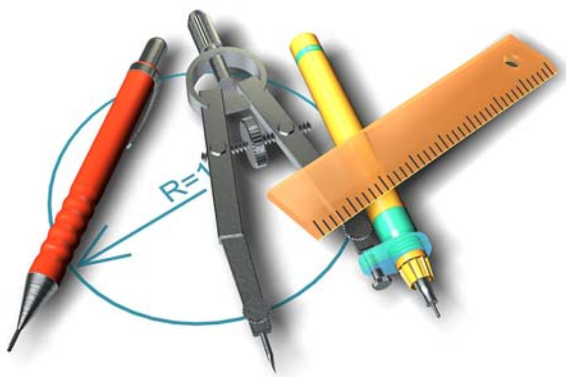


TABLE 8

The Nation's Report Card 2003 State Assessment

Total number of students assessed, percentage of students sampled that were excluded, and average mathematics scale scores, grade 8 public schools: By state, 2003.

	Grade 8		
	Number assessed	Percentage excluded	Average scale score
Alabama	2,563	2 (0.4)	262 (1.5)
Alaska	2,545	1 (0.2)	279 (0.9)
Arizona	2,713	4 (0.5)	271 (1.2)
Arkansas	2,582	2 (0.4)	266 (1.2)
California	5,512	3 (0.4)	267 (1.2)
Colorado	2,757	2 (0.3)	283 (1.1)
Connecticut	2,698	4 (0.4)	284 (1.2)
Delaware	2,455	9 (0.6)	277 (0.7)
Florida	2,483	3 (0.5)	271 (1.5)
Georgia	4,246	2 (0.3)	270 (1.2)
Hawaii	2,824	4 (0.4)	266 (0.8)
Idaho	2,708	1 (0.2)	280 (0.9)
Illinois	4,122	4 (0.5)	277 (1.2)
Indiana	2,656	2 (0.3)	281 (1.1)
Iowa	2,932	2 (0.3)	284 (0.8)
Kansas	2,934	3 (0.4)	284 (1.3)
Kentucky	2,833	4 (0.7)	274 (1.2)
Louisiana	2,370	5 (0.6)	266 (1.5)
Maine	2,861	4 (0.4)	282 (0.9)
Maryland	2,406	4 (0.7)	278 (1.0)
Massachusetts	3,773	3 (0.6)	287 (0.9)
Michigan	2,652	5 (0.6)	276 (2.0)
Minnesota	2,645	2 (0.3)	291 (1.1)
Mississippi	2,625	5 (0.5)	261 (1.1)
Missouri	2,735	4 (0.6)	279 (1.1)
Montana	2,643	2 (0.3)	286 (0.8)
Nebraska	2,469	4 (0.4)	282 (0.9)
Nevada	2,646	2 (0.2)	268 (0.8)
New Hampshire	2,829	3 (0.5)	286 (0.8)
New Jersey	2,810	2 (0.4)	281 (1.1)
New Mexico	3,217	2 (0.4)	263 (1.0)
New York	3,422	5 (0.7)	280 (1.1)
North Carolina	4,093	4 (0.5)	281 (1.0)
North Dakota	2,684	1 (0.2)	287 (0.8)
Ohio	3,523	5 (0.8)	282 (1.3)
Oklahoma	2,855	2 (0.3)	272 (1.1)
Oregon	2,671	3 (0.4)	281 (1.3)
Pennsylvania	2,776	2 (0.3)	279 (1.1)
Rhode Island	2,669	4 (0.3)	272 (0.7)
South Carolina	2,471	7 (0.8)	277 (1.3)
South Dakota	2,839	2 (0.3)	285 (0.8)
Tennessee	2,610	3 (0.4)	268 (1.8)
Texas	4,398	7 (0.6)	277 (1.1)
Utah	2,726	3 (0.4)	281 (1.0)
Vermont	2,650	3 (0.4)	286 (0.8)
Virginia	2,776	7 (0.6)	282 (1.3)
Washington	2,629	2 (0.3)	281 (0.9)
West Virginia	2,365	3 (0.4)	271 (1.2)
Wisconsin	2,591	3 (0.4)	284 (1.3)
Wyoming	2,720	1 (0.2)	284 (0.7)
DC	1,888	6 (0.4)	243 (0.8)
DoDEA/DDESS	709	2 (0.5)	282 (1.5)
DoDEA/DoDDS	2,256	1 (0.2)	286 (0.7)

NOTE: The NAEP mathematics scale ranges from 0 to 500. The standard errors of the statistics in the table appear in parentheses.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Mathematics Assessment.

Appendix
Overview of Procedures Used for the NAEP 2003 Mathematics Assessment

The NAEP 2003 Mathematics Assessment

The National Assessment Governing Board (NAGB), created by Congress in 1988, is responsible for formulating policy for NAEP. NAGB is specifically charged with developing assessment objectives and test specifications. The mathematics framework used for the 2003 assessment had its origins in a framework developed for the 1990 mathematics assessment under contract with the Council of Chief State School Officers (CCSSO). The CCSSO project considered objectives and frameworks for mathematics instruction at the state, district, and school levels. The project also examined curricular frameworks on which previous NAEP assessments were based, consulted with leaders in mathematics education, and considered a draft version of the National Council of Teachers of Mathematics (NCTM) *Curriculum and Evaluation Standards for School Mathematics*.[1] This project resulted in a "content-by-ability" matrix design used to guide both the NAEP 1990 and 1992 mathematics assessments. The design was reported in *Mathematics Objectives: 1990 Assessment*.[2]

Prior to 1990, mathematics was assessed based on an earlier framework, which also was used to develop NAEP long-term trend assessments. Because the long-term trend assessments all use the same test booklets, it is possible to compare students' performance across many assessment years. However, the NAEP main mathematics assessment that was administered in 2003 is comparable only to the other assessments based on the 1990 framework—1990, 1992,1996, and 2000.

The 1996 assessment was based on the first update of the 1990 NAEP mathematics framework since the release of the NCTM *Curriculum and Evaluation Standards for School Mathematics* in 1989.[3] This update was conducted by the College Board and reflected refinements in the earlier framework specifications while ensuring comparability of results across the 1990, 1992, and 1996 assessments. Since the 2003 framework is the same as the 1996 update, the assessment results from 1990 to 2003 can be compared. The refinements that distinguish the framework used in the 1996, 2000, and 2003 assessments from the assessments conducted in 1990 and 1992 include the following:

- moving away from the rigid content-by-ability matrix (forcing items to be classified in cells of a matrix limited the possibility of assessing students' ability to reason in rich problem-solving situations and to make connections among the content areas);
- including the three achievement levels—*Basic*, *Proficient*, and *Advanced*—described in this report;
- allowing individual questions to be classified in more than one content area (since the option to classify questions in more than one content area provides greater opportunity to measure student ability in content settings that more closely approximate real-world situations);
- including the mathematics ability categories (conceptual understanding, procedural understanding, and problem solving) as well as the process goals (communication and connections) from the NCTM *Standards*;
- including more constructed-response questions in the 1996, 2000, and 2003 assessments than were included in 1990 and 1992; and
- revisiting some of the content areas to make sure they reflect recent curricular emphases.

The following figure describes the five content areas that constitute the NAEP mathematics assessment. These content areas apply to each of the three grades assessed by NAEP. The questions designed to test the various content areas at a particular grade level tend to reflect the expectations normally associated with instruction at that grade level.

	Descriptions of the Five NAEP Mathematics Content Areas
Number Sense, Properties, and Operations	This content area focuses on students' understanding of numbers (whole numbers, fractions, decimals, integers, real numbers, and complex numbers), operations, and estimation, and their application to real-world situations. At grade 4, the emphasis is on the development of number sense through connecting various models to their numerical representations, and an understanding of the meaning of addition, subtraction, multiplication, and division. At grade 8, number sense is extended to include positive and negative numbers, as well as properties and operations involving whole numbers, fractions, decimals, integers, and rational numbers.
Measurement	This content area focuses on an understanding of the process of measurement and the use of numbers and measures to describe and compare mathematical and real-world objects. Students are asked to identify attributes, select appropriate units and tools, apply measurement concepts, and communicate measurement-related ideas. At grade 4, the focus is on time, money, temperature, length, perimeter, area, capacity, weight/mass, and angle measure. At grade 8, this content area includes these measurement concepts, but the focus shifts to more complex measurement problems that involve volume or surface area or that require students to combine shapes and to translate and apply measures. Eighth-grade students also solve problems involving proportional thinking (such as scale drawing or map reading) and do applications that involve the use of complex measurement formulas.
Geometry and Spatial Sense	This content area is designed to extend beyond low-level identification of geometric shapes to include transformations and combinations of those shapes. Informal constructions and demonstrations (including drawing representations) along with their justifications take precedence over more traditional types of compass-and-straightedge constructions and proofs. At grade 4, students are asked to model properties of shapes under simple combinations and transformations, and to use mathematical communication skills to draw figures from verbal descriptions. At grade 8, students are asked to expand their understanding to include properties of angles and polygons. They are also asked to apply reasoning skills to make and validate conjectures about transformations and combinations of shapes.
Data Analysis, Statistics, and Probability	This content area emphasizes the appropriate methods for gathering data, the visual exploration of data, various ways of representing data, and the development and evaluation of arguments based on data analysis. At grade 4, students are asked to apply their understanding of numbers and quantities by solving problems that involve data. Fourth-graders are asked to interact with a variety of graphs, to make predictions from data and explain their reasoning, to deal informally with measures of central tendency, and to use the basic concepts of chance in meaningful contexts. At grade 8, students are asked to analyze statistical claims and to design experiments, and they are asked to use simulations to model real-world situations. This content area focuses on eighth-graders' basic understanding of sampling, their ability to make predictions based on experiments or data, and their ability to use some formal terminology related to probability, data analysis, and statistics.
Algebra and Functions	This content area extends from work with simple patterns at grade 4 to basic algebra concepts at grade 8. The grade 4 assessment involves informal demonstration of students' abilities to generalize from patterns, including the justification of their generalizations. Students are expected to translate between mathematical representations, to use simple equations, and to do basic graphing. At grade 8, the assessment includes more algebraic notation, stressing the meaning of variables and an informal understanding of the use of symbolic representations in problem-solving contexts. Students are asked to use variables to represent a rule underlying a pattern. Eighth-graders are asked to demonstrate a beginning understanding of equations and functions and the ability to solve simple equations and inequalities.

The assessment framework specifies not only the particular areas that should be assessed, but also the percentage of the assessment questions that should be devoted to each of the content areas. The target percentage distribution for content areas as specified in the framework is presented in the following table. The distribution of items among the content areas is a critical feature of the assessment design, since it reflects the relative importance and value given to each. The target percentages at eighth grade differ from those at fourth grade because of a shift in curricular emphasis. For example, in grade 4 there is more emphasis on number sense, properties, and operations than on algebra and functions. In grade 8, the percentage of algebra and functions items increases, and the percentage of number sense, properties, and operations items decreases. The actual content of the assessment is close to the targeted distribution.

Target percentage distribution of items, by content area and grade: 1990–2003

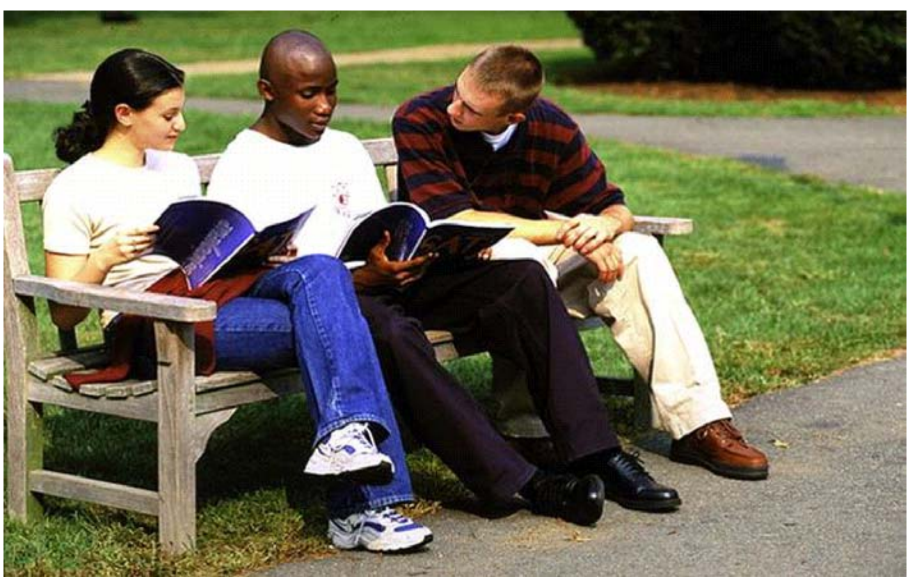
	Grade 4		Grade 8	
	1990–1992	1996–2003	1990–1992	1996–2003
Number sense, properties, and operations	45	40	30	25
Measurement	20	20	15	15
Geometry and spatial sense	15	15	20	20
Data analysis, statistics, and probability	10	10	15	15
Algebra and functions	10	15	20	25

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1990, 1992, 1996, 2000, and 2003 Mathematics Assessments.

1. National Council of Teachers of Mathematics. (1989). *Curriculum and Evaluation Standards for School Mathematics.* Reston, VA: Author.

2. National Assessment of Educational Progress. (1988). *Mathematics Objectives: 1990 Assessment*. Princeton, NJ: Author.

3. National Assessment Governing Board. *Mathematics Framework for the 1996 National Assessment of Educational Progress*. Washington, DC: Author.



The Assessment Design

Each student who participated in the NAEP 2003 mathematics assessment received a booklet containing four sections: two sets of cognitive questions, a set of general background questions, and a set of subject-specific background questions. Assessments for each grade consisted of 10 sets of cognitive questions or "blocks." Some items from the 1990, 1992, 1996, and 2000 assessments were carried forward to 2003 to allow for the measurement of trends across time. Two new blocks were developed for the 2003 assessment as specified by the updated framework.

Three types of questions are used in the assessment: multiple-choice, short constructed-response, and extended constructed-response. The following table shows the distribution of questions administered from 1990 to 2003 by type for each grade level. The total number of questions administered has varied somewhat across the assessment years due to the inclusion of special study blocks in certain years. The number of questions used in the main scaling, however, has remained relatively consistent.

Distribution of questions administered, by question type and grade: 1990–2003

	Grade 4					Grade 8				
	1990	1992	1996	2000	2003	1990	1992	1996	2000	2003
Multiple-choice	102	99	81	87	114	149	118	102	100	129
Short constructed-response	41	59	64	50	60	42	65	69	51	58
Extended constructed-response	---	5	13	8	8	---	6	12	9	10
Total	143	163	158	145	182	191	189	183	160	197

--- No extended constructed-response questions were included in the 1990 assessment.
NOTE: Short constructed-response questions included in the 1990 and 1992 assessments were scored dichotomously.
New short constructed-response questions included in the 1996, 2000, and 2003 assessments were scored to allow for partial credit.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1990, 1992, 1996, 2000, and 2003 Mathematics Assessments.

The assessment design allowed maximum coverage of mathematics abilities at each grade, while minimizing the time burden for any one student. This was accomplished through the use of matrix sampling of items in which representative samples of students took various portions of the entire pool of assessment questions. Individual students are required to take only a small portion of the assessment, but the aggregate results across the entire assessment allow broad reporting of mathematics abilities for the targeted population.

In addition to matrix sampling, the assessment design utilized a procedure for distributing blocks across booklets that controlled for position and context effects. Students received different blocks of questions in their booklets according to a procedure called "balanced incomplete block (BIB) spiraling." This procedure assigned blocks of questions so that every block appeared in each of the two possible positions within a booklet an equal number of times. Also, every block of questions was paired with every other block. The spiraling aspect of this procedure cycles the booklets for administration so that, typically, only a few students in any assessment session receive the same booklet.

In addition to the student assessment booklets, three other instruments provided data relating to the assessment: a teacher questionnaire, a school questionnaire, and a questionnaire for students with disabilities (SD) and limited-English-proficient students (LEP). The teacher questionnaire was administered to the mathematics teachers of the fourth- and eighth-grade students participating in the assessment. The questionnaire took approximately 20 minutes to complete and focused on the teacher's general background and experience, the teacher's background related to mathematics, and classroom information about mathematics instruction.

The school questionnaire was given to the principal or other administrator in each participating school and took about 20 minutes to complete. The questions asked about school policies, programs, facilities, and the demographic composition and background of the students and teachers at the school.

The SD/LEP questionnaire was completed by a school staff member knowledgeable about those students selected to participate in the assessment who were identified as having an Individualized Education Program (IEP) or equivalent plan (for reasons other than being gifted or talented) or having limited English proficiency. An SD/LEP questionnaire was completed for each identified student regardless of whether the student participated in the assessment. Each SD/LEP questionnaire took approximately three minutes to complete and asked about the student and the special-education programs in which he or she participated.

Data Collection and Scoring

The NAEP 2003 mathematics assessment was conducted from January to March 2003 by contractors to the U.S. Department of Education. Trained field staff from Westat conducted the data collection. Materials from the 2003 assessment were shipped to Pearson, where trained staff evaluated the responses to the constructed-response questions using scoring rubrics or guides prepared by Educational Testing Service (ETS). Each constructed-response question had a unique scoring guide that defined the criteria used to evaluate students' responses. The extended constructed-response questions were evaluated with four- and five-level guides, and many of the short constructed-response questions were rated according to three-level guides that permitted partial credit. Other short constructed-response questions were scored as either correct or incorrect.

For the 2003 mathematics assessment, 4,719,464 constructed responses were scored. This number includes rescoring to monitor interrater reliability. The within-year average percentage of exact agreement for the 2003 national reliability sample was 95 percent at both fourth and eighth grade.

Data Analysis and IRT Scaling

After the professional scoring, all information was transcribed into the NAEP database at ETS. Each processing activity was conducted with rigorous quality control. After the assessment information was compiled in the database, the data were weighted according to the population structure. The weighting for the national and state samples reflected the probability of selection for each student as a result of the sampling design, adjusted for nonresponse.[1]

Analyses were then conducted to determine the percentages of students who gave various responses to each cognitive and background question. In determining these percentages for the cognitive questions, a distinction was made between missing responses at the end of a block (i.e., missing responses after the last question the student answered) and missing responses before the last observed response. Missing responses before the last observed response were considered intentional omissions. In analysis, omitted responses to multiple-choice items were scored as fractionally correct.[2] Omitted responses for constructed-response items were placed into the lowest score category. Missing responses after the last observed response were considered "not reached" and treated as if the questions had not been presented to the student. In calculating response percentages for each question, only students classified as having been presented the question were included in the denominator of the statistic.

It is standard NAEP practice to treat all nonrespondents to the last question in a block as if they had not reached the question. For multiple-choice and short constructed-response questions, this practice produces a reasonable pattern of results in that the proportion reaching the last question is not dramatically smaller than the proportion reaching the next-to-last question. However, for mathematics blocks that ended with extended constructed-response questions, there may be extremely large drops in the proportion of students attempting some of the final questions. Therefore, for blocks ending with an extended constructed-response question, students who answered the next-to-last question but did not respond to the extended constructed-response question were classified as having intentionally omitted the last question.

Item Response Theory (IRT) was used to estimate average mathematics scale scores for the nation and for various subgroups of interest within the nation. IRT models the probability of answering a question in a certain way as a mathematical function of proficiency or skill. The main purpose of IRT analysis is to provide a common scale on which performance can be compared among groups such as those defined by characteristics, including gender and race/ethnicity, even when students receive different blocks of items. One desirable feature of IRT is that it locates items and students on this common scale. In contrast to classical test theory, IRT does not rely solely on the total number of correct item responses, but uses the particular patterns of student responses to items in determining the student location on the scale. As a result, adding items that function at a particular point on the scale to the assessment does not change the location of the students on the scale, even though students may respond correctly to more items. It does increase the relative precision with which students are measured, particularly those students whose scale locations are close to the additional items.

The results for 1990, 1992, 1996, 2000, and 2003 are presented on the NAEP mathematics composite scale. For the NAEP mathematics assessment, a scale ranging from 0 to 500 was used to report performance in each of the five mathematics content areas at each grade: number sense, properties, and operations; measurement; geometry and spatial sense; data analysis, statistics, and probability; and algebra and functions. The scales summarize student performance across all three types of questions in the assessment (multiple-choice, short constructed-response, and extended constructed-response).

In producing the mathematics scales, three distinct IRT models were used. Multiple-choice questions were scaled using the three-parameter logistic (3PL) model; short constructed-response questions rated as acceptable or unacceptable were scaled using the two-parameter logistic (2PL) model; and short constructed-response questions rated according to a three-level guide, as well as extended constructed-response questions rated on a four- or five-level guide, were scaled using a generalized partial-credit (GPC) model.[3] Developed by ETS and first used in 1992, the GPC model permits the scaling of questions scored according to multipoint rating schemes. The model takes full advantage of the information available from each of the student response categories used for these more complex constructed-response questions.

The mathematics scale is composed of three types of questions: multiple-choice, short constructed-response (scored either dichotomously or allowing for partial credit), and extended constructed-response (scored according to a partial-credit model). Unfortunately, the question of how much information different types of questions contribute to the mathematics scale has no simple answer. The information provided by a given question is determined by the IRT model used to scale the question. It is a function of the item parameters and varies by level of mathematics proficiency.[4] Thus, the answer to the query "How much information do the different types of questions provide?" will differ for each level of mathematics performance. When considering the composite mathematics scale, the answer is even more complicated. The mathematics data are scaled separately by the content areas. The composite scale is a weighted combination of these subscales. IRT information functions are only strictly comparable when they are derived from the same calibration. Because the composite scale is based on five separate calibrations, there is no direct way to compare the information provided by the questions on the composite scale.

Because the NAEP design gives each student a small proportion of the pool of assessment items, the assessment cannot provide reliable information about individual performance. Traditional test scores for individual students, even those based on IRT, would result in misleading estimates of population characteristics, such as subgroup means and percentages of students at or above a certain scale-score level. However, it is NAEP's goal to estimate these population characteristics. NAEP's objectives can be achieved with methodologies that produce estimates of the population-level parameters directly, without the intermediary computation of estimates of individuals.[5] This is accomplished using marginal estimation scaling model techniques for latent variables. Under the assumptions of the scaling models, these population estimates will be consistent in the sense that the estimates approach the model-based population values as the sample size increases. This would not be the case for population estimates obtained by aggregating optimal estimates of individual performance.[6]

---

1. Weighting procedures are described more fully under the topic "Weighting and Variance Estimation."

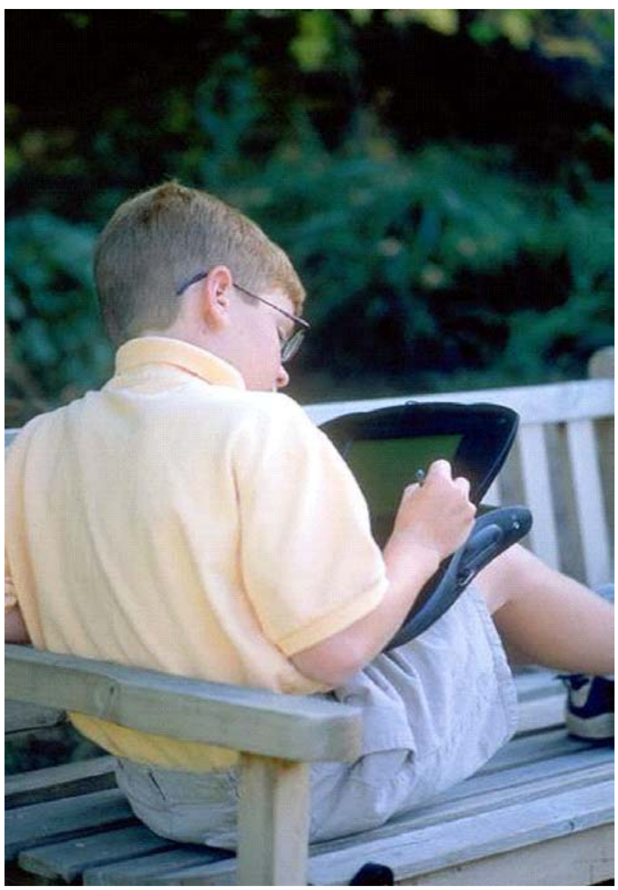
2. Lord, F. M. (1980). *Applications of Item Response Theory to Practical Testing Problems,* p. 229. Hillsdale, NJ: Lawrence Erlbaum Associates.

3. Muraki, E. (1992). A Generalized Partial Credit Model: Application of an EM Algorithm. *Applied Psychological Measurement, 16*(2), 159–176.

4. Donoghue, J. R. (1994). An Empirical Examination of the IRT Information of Polytomously Scored Reading Items Under the Generalized Partial Credit Model. *Journal of Educational Measurement, 31*(4), 295–311.

5. Mislevy, R. J., and Sheehan, K. M. (1987). Marginal Estimation Procedures. In A. E. Beaton (Ed.), *Implementing the New Design: The NAEP 1983–1984 Technical Report* (Report No. 15-TR-20), pp. 260–293. Princeton, NJ: Educational Testing Service.

6. For theoretical and empirical justification of the procedures employed, see Mislevy, R. J. (1988). Randomization Based Inferences About Latent Variables from Complex Samples. *Psychometrika, 56*(2), 177–196. For additional discussion, see Thomas, N. (1993). Asymptotic Corrections for Multivariate Posterior Moments with Factored Likelihood Functions. *Journal of Computational and Graphical Statistics*, *25*, 351-372. Also see Mazzeo, J., Donoghue, J. R., and Johnson, M. (under review). Marginal Estimation in NAEP: Current Operational Procedures and *AM*.



Weighting and Variance Estimation

A complex sampling design was used to select the students who were assessed. The properties of a sample selected through such a design can be very different from those of a simple random sample, in which every student in the target population has an equal chance of selection and in which the observations from different sampled students can be considered to be statistically independent of one another. Therefore, the properties of the sample for the data collection design were taken into account during the analysis of the assessment data.

One way that the properties of the sample design were addressed was by using sampling weights to account for the fact that the probabilities of selection were not identical for all students. All population and subpopulation characteristics based on the assessment data were estimated using sampling weights. These weights included adjustments for school and student nonresponse.

Prior to 2003, the national samples used weights that had been poststratified to the Census or Current Population Survey (CPS) totals for the populations being assessed. Due to concerns about the availability of appropriate targets for poststratification as a result of changes in the reporting of race in the 2000 census, nonpoststratified weights have been used in the analysis of national samples since 2003. The state NAEP samples have always been analyzed using nonpoststratified weights, since there were no targets available from CPS to use in poststratification.

Not only must appropriate estimates of population characteristics be derived, but appropriate measures of the degree of uncertainty must be obtained for those statistics. Two components of uncertainty are accounted for in the variability of statistics based on student ability: 1) the uncertainty due to sampling only a relatively small number of students, and 2) the uncertainty due to sampling only a portion of the cognitive domain of interest. The first component accounts for the variability associated with the estimated percentages of students who had certain background characteristics or who answered a certain cognitive question correctly.

Because NAEP uses complex sampling procedures, conventional formulas for estimating sampling variability that assume simple random sampling are inappropriate. NAEP uses a jackknife replication procedure to estimate standard errors. The jackknife standard error provides a reasonable measure of uncertainty for any student information that can be observed without error. However, because each student typically responds to only a few questions within any mathematics content area, the scale score for any single student would be imprecise. In this case, NAEP's marginal estimation methodology can be used to describe the performance of groups and subgroups of students. The estimate of the variance of the students' posterior scale score distributions (which reflect the imprecision due to lack of measurement accuracy) is computed. This component of variability is then included in the standard errors of NAEP scale scores.[1]

Typically, when the standard error is based on a small number of students or when the group of students is enrolled in a small number of schools, the amount of uncertainty associated with the estimation of standard errors may be quite large. Estimates of standard errors subject to a large degree of uncertainty are followed on the tables by the "!" symbol to indicate that the nature of the sample does not allow accurate determination of the variability of the statistic. In such cases, the standard errors—and any confidence intervals or significance tests involving these standard errors—should be interpreted cautiously.

The reader is reminded that, as with findings from all surveys, NAEP results are subject to other kinds of error, including the effects of imperfect adjustment for student and school nonresponse and unknowable effects associated with the particular instrumentation and data collection methods. Nonsampling errors can be attributed to a number of sources—inability to obtain complete information about all selected schools in the sample (some students or schools refused to participate, or students participated but answered only certain questions); ambiguous definitions; differences in interpreting questions; inability or unwillingness to give correct background information; mistakes in recording, coding, or scoring data; and other errors in collecting, processing, sampling, and estimating missing data. The extent of nonsampling errors is difficult to estimate and, because of their nature, the impact of such errors cannot be reflected in the data-based estimates of uncertainty provided in NAEP reports.

---

1. For further details, see Johnson, E. G., and Rust, K. F. (1992). Population Inferences and Variance Estimation for NAEP Data. *Journal of Educational Statistics, 17*(2), 175–190.

Drawing Inferences from the Results

The reported statistics are estimates and are therefore subject to a measure of uncertainty. There are two sources of such uncertainty. First, NAEP uses a sample of students rather than testing all students. Second, all assessments have some amount of uncertainty related to the fact that they cannot ask all questions that might be asked in a content area. The magnitude of this uncertainty is reflected in the standard error of each of the estimates. When the percentages or average scale scores of certain groups are compared, the estimated standard error should be taken into account. Therefore, the comparisons are based on statistical tests that consider the estimated standard errors of those statistics and the magnitude of the difference among the averages or percentages.

For the data in this report, all the estimates have corresponding estimated standard errors of the estimates. For example, the following tables show the average national public-school scale score for the NAEP 1990–2003 national assessments and the percentage of students within each achievement-level range and at or above achievement levels. In both tables, estimated standard errors appear in parentheses next to each estimated scale score or percentage. For the estimated standard errors corresponding to other data from this report, the reader can go to the Data Tool on the NCES web site at http://nces.ed.gov/nationsreportcard/naepdata/.

Average mathematics scale scores and standard errors, grades 4 and 8: 1990–2003

	Accommodations not permitted				Accommodations permitted		
	1990	**1992**	**1996**	**2000**	**1996**	**2000**	**2003**
Grade 4	213(0.9)*	220(0.7)*	224(0.9)*	228(0.9)*	224(1.0)*	226(0.9)*	235(0.2)
Grade 8	263(1.3)*	268(0.9)*	272(1.1)*	275(0.8)*	270(0.9)*	273(0.8)*	278(0.3)

* Significantly different from 2003.
NOTE: Standard errors of the estimated scale scores appear in parentheses.
In addition to allowing for accommodations, the accommodations-permitted results (1996–2003) differ slightly from previous years' results, and from previous reported results for 1996 and 2000 due to changes in sample weighting procedures.
Significance tests were performed using unrounded numbers.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1990, 1992, 1996, 2000, and 2003 Mathematics Assessments.

Percentage of students and standard errors, by mathematics achievement level, grades 4 and 8: 1990–2003

		Below *Basic*	At *Basic*	At or above *Basic*	At or above *Proficient*
Grade 4					
Accommodations not permitted	1990	50(1.4)*	1(0.4)*	50(1.4)*	13(1.2)*
	1992	41(1.0)*	2(0.3)*	59(1.0)*	18(1.0)*
	1996	36(1.2)*	2(0.3)*	64(1.2)*	21(0.9)*
	2000	31(1.1)*	3(0.3)*	69(1.1)*	26(1.1)*
Accommodations permitted	1996	37(1.3)*	2(0.3)*	63(1.3)*	21(1.1)*
	2000	35(1.3)*	3(0.3)*	65(1.3)*	24(1.0)*
	2003	23(0.3)	4(0.1)	77(0.3)	32(0.3)
Grade 8					
Accommodations not permitted	1990	48(1.4)*	2(0.3)*	52(1.4)*	15(1.1)*
	1992	42(1.1)*	3(0.4)*	58(1.1)*	21(1.0)*
	1996	38(1.1)*	4(0.5)*	62(1.1)*	24(1.1)*
	2000	34(0.8)*	5(0.5)	66(0.8)*	27(0.9)
Accommodations permitted	1996	39(1.0)*	4(0.4)*	61(1.0)*	23(1.0)*
	2000	37(0.9)*	5(0.4)	63(0.9)*	26(0.8)*
	2003	32(0.3)	5(0.1)	68(0.3)	29(0.3)

* Significantly different from 2003.
NOTE: Standard errors of the estimated percentages appear in parentheses.
Detail may not sum to totals because of rounding.
In addition to allowing for accommodations, the accommodations-permitted results (1996–2003) differ slightly from previous years' results, and from previously reported results for 1996 and 2000 due to changes in sample weighting procedures.
Significance tests were performed using unrounded numbers.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1990, 1992, 1996, 2000, and 2003 Mathematics Assessments.

Using confidence intervals based on the standard errors provides a way to take into account the uncertainty associated with sample estimates and to make inferences about the population averages and percentages in a manner that reflects that uncertainty. An estimated sample average scale score plus or minus 1.96 standard errors approximates a 95 percent confidence interval for the corresponding population quantity. This statement means that one can conclude with an approximately 95 percent level of confidence that the average performance of the entire population of interest (e.g., all fourth-grade students in public schools) is within plus or minus 1.96 standard errors of the sample average.

For example, suppose that the average mathematics scale score of the students in a particular group was 256 with an estimated standard error of 1.2. An approximately 95 percent confidence interval for the population quantity would be as follows:

Average ± 1.96 standard errors
256 ± 1.96 x 1.2
256 ± 2.4
(253.6, 258.4)

Thus, one can conclude with a 95 percent level of confidence that the average scale score for the entire population of students in that group is between 253.6 and 258.4. It should be noted that this example and the examples in the following sections are illustrative. More precise estimates carried out to one or more decimal places are used in the actual analyses.

Similar confidence intervals can be constructed for percentages, if the percentages are not extremely large or extremely small. Extreme percentages should be interpreted with caution. Adding or subtracting the standard errors associated with extreme percentages could cause the confidence interval to exceed 100 percent or fall below 0 percent, resulting in numbers that are not meaningful.

Analyzing Group Differences in Averages and Percentages

Statistical tests determine whether, based on the data from the groups in the sample, there is strong enough evidence to conclude that the averages or percentages are actually different for those groups in the population. If the evidence is strong (i.e., the difference is statistically significant), the report describes the group averages or percentages as being different (e.g., one group performed higher or lower than another group), regardless of whether the sample averages or percentages appear to be approximately the same. The reader is cautioned to rely on the results of the statistical tests rather than on the apparent magnitude of the difference between sample averages or percentages when determining whether the sample differences are likely to represent actual differences among the groups in the population.

To determine whether a real difference exists between the average scale scores (or percentages of a certain attribute) for two groups in the population, one needs to obtain an estimate of the degree of uncertainty associated with the difference between the averages (or percentages) of these groups for the sample. This estimate of the degree of uncertainty, called the "standard error of the difference" between the groups, is obtained by taking the square of each group's standard error, summing the squared standard errors, and taking the square root of that sum.

Standard Error of the Difference = $SE_{A-B} = \sqrt{(SE_A^2 + SE_B^2)}$

The standard error of the difference can be used, just like the standard error for an individual group average or percentage, to help determine whether differences among groups in the population are real. The difference between the averages or percentages of the two groups plus or minus 1.96 standard errors of the difference represents an approximately 95 percent confidence interval. If the resulting interval includes zero, there is insufficient evidence to claim a real difference between the groups in the population. If the interval does not contain zero, the difference between the groups is statistically significant at the 0.05 level.

The following example of comparing groups addresses the problem of determining whether the average mathematics scale score of group A is higher than that of group B. The sample estimates of the average scale scores and estimated standard errors are as follows:

Group	Average Scale Score	Standard Error
A	218	0.9
B	216	1.1

The difference between the estimates of the average scale scores of groups A and B is two points (218–216). The standard error of this difference is $\sqrt{(0.9^2 + 1.1^2)} = 1.4$

Thus, an approximately 95 percent confidence interval for this difference is plus or minus 1.96 standard errors of the difference:

$$2 \pm 1.96 \times 1.4$$
$$2 \pm 2.7$$
$$(-0.7, 4.7)$$

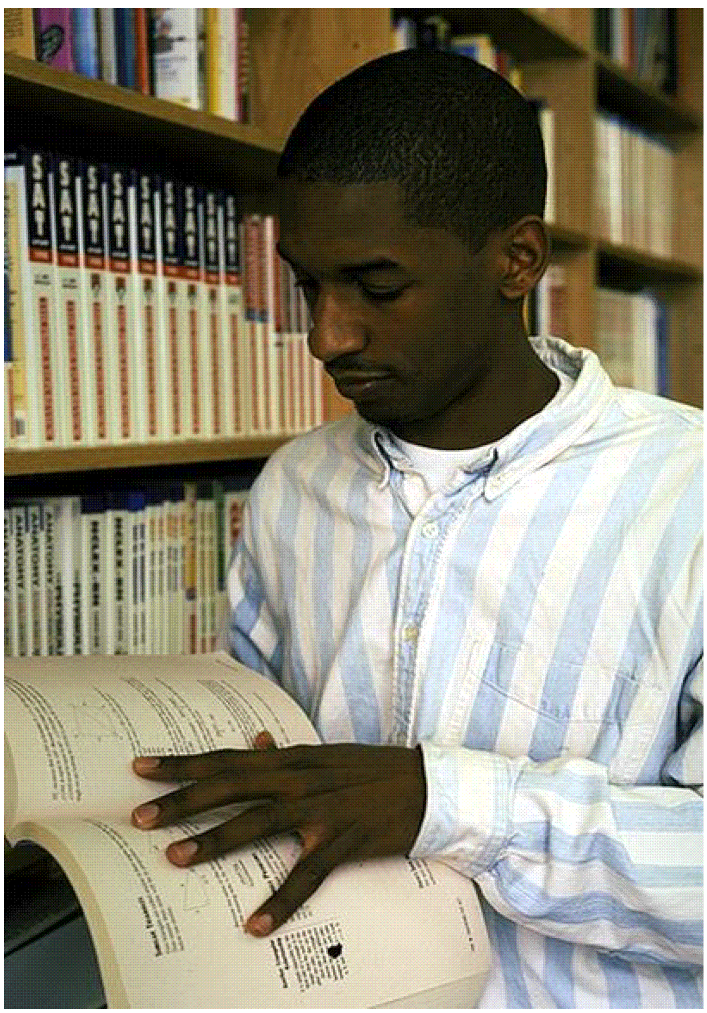
The value zero is within the confidence interval; therefore, there is insufficient evidence to conclude that group A outperformed group B.

The procedure above is appropriate to use when it is reasonable to assume that the groups being compared have been independently sampled for the assessment. Such an assumption is clearly warranted when comparing results across assessment years (e.g., comparing the 2000 and 2003 results for a particular state or subgroup) or when comparing results for one state with another. This is the approach used for NAEP reports when comparisons involving independent groups are made. The assumption of independence is violated to some degree when comparing group results for the nation or a particular state (e.g., comparing national 2003 results for males and females), since these samples of students have been drawn from the same schools. When the groups being compared do not share students (as is the case, for example, comparing males and females) the impact of this violation of the independence assumption on the outcome of the statistical tests is assumed to be small, and NAEP, by convention, has, for computational convenience, routinely applied the procedures described above to those cases as well.

When making comparisons of results for groups that share a considerable proportion of students in common, it is not appropriate to ignore such dependencies. In such cases, NAEP has used procedures appropriate to comparing dependent groups. When the dependence in group results is due to the overlap in samples (e.g., when a subgroup is being compared to a total group), a simple modification of the usual standard error of the difference formula can be used. The formula for such cases is $SE_{Total-Subgroup} = \sqrt{(SE^2_{Total} + SE^2_{Subgroup} - 2pSE^2_{Subgroup})}$

where p is the proportion of the total group contained in the subgroup.[1] This formula was used for this report when a state was compared to the aggregate nation.

1. This is a special form of the common formula for standard error of dependent samples. The standard formula can be found, for example, in Kish, L. (1995). *Survey Sampling.* New York: John Wiley and Sons, Inc.



Conducting Multiple Tests

The procedures used to determine whether group differences in the samples represent actual differences among the groups in the population and the certainty ascribed to intervals (e.g., a 95 percent confidence interval) are based on statistical theory that assumes that only one confidence interval or test of statistical significance is being performed. However, there are times when many different groups are being compared (i.e., multiple sets of confidence intervals are being analyzed). In sets of confidence intervals, statistical theory indicates that the certainty associated with the entire set of intervals is less than that attributable to each individual comparison from the set. To hold the significance level for the set of comparisons at a particular level (e.g., 0.05), the standard methods must be adjusted by multiple comparison procedures.[1] One such procedure, the Benjamini-Hochberg False Discovery Rate (FDR) procedure, was used to control the certainty level.[2]

Unlike other multiple comparison procedures that control the familywise error rate (i.e., the probability of making even one false rejection in the set of comparisons), the FDR procedure controls the expected proportion of falsely rejected hypotheses. Furthermore, the FDR procedure used in NAEP is considered appropriately less conservative than familywise procedures for large families of comparisons.[3] Therefore, the FDR procedure is more suitable for multiple comparisons in NAEP than other procedures.

To illustrate how the FDR procedure is used, consider the comparisons of current and previous years' average scale scores for the five groups presented in the following table. The test statistic shown is the difference in average scale scores divided by the estimated standard error of the difference. (Rounding of the data occurs after the test is done.)

Example of False Discovery Rate comparisons of average scale scores for different groups of students

	Previous year		Current year		Previous year and current year			
	Average scale score	Standard error	Average scale score	Standard error	Difference in averages	Standard error of difference	Test Statistic	Percent confidence*
Group 1	224	1.3	226	1.0	2.08	1.62	1.29	20
Group 2	187	1.7	193	1.7	6.31	2.36	2.68	1
Group 3	191	2.6	197	1.7	6.63	3.08	2.15	4
Group 4	229	4.4	232	4.6	3.24	6.35	0.51	62
Group 5	201	3.4	196	4.7	-5.51	5.81	-0.95	35

* The percent confidence is 2(1-F(x)), where F(x) is the cumulative distribution of the t-distribution with the degrees of freedom adjusted to reflect the complexities of the sample design.

The difference in average scale scores and its estimated standard error can be used to find an approximately 95 percent confidence interval or they can be used to identify a confidence percentage. The confidence percentage for the test statistics is identified from statistical tables. The significance level from the statistical tables can be directly compared to 100 - 95 = 5 percent.

If the comparison of average scale scores across two years was made for only one of the five groups, there would be a significant difference between the average scale scores for the two years at a significance level of less than 5 percent. However, because we are interested in the difference in average scale scores across the two years for all five of the groups, comparing each of the significance levels to 5 percent is not adequate. Groups of students defined by shared characteristics, such as racial/ethnic groups, are treated as sets or families when making comparisons. However, comparisons of average scale scores for each pair of years were treated separately, so the steps described in this example would be replicated for the comparison of other current and previous year average scale scores.

Using the FDR procedure to take into account that all comparisons are of interest to us, the percents of confidence in the example are ordered from largest to smallest: 62, 35, 20, 4, and 1. In the FDR procedure, 62 percent confidence for the group 4 comparison would be compared to 5 percent; 35 percent for the group 5 comparison would be compared to 0.05 x (5-1)/5=0.04=4 percent;[4] 20 percent for the group 1 comparison would be compared to 0.05 x (5-2)/5=0.03=3 percent; 4 percent for the group 3 comparison would be compared to 0.05 x (5-3)/5=0.02=2 percent; and 1 percent for the group 2 comparison (actually slightly smaller than 1 prior to rounding) would be compared to 0.05 x (5-4)/5=0.01=1 percent. The procedure stops with the first contrast found to be significant. The last of these comparisons is the only one for which the percent confidence is smaller than the FDR procedure value. The difference between the current year's and previous years' average scale scores for the group 2 students is significant; for all of the other groups, average scale scores for current and previous years are not significantly different from one another. In practice, a very small number of counterintuitive results occur when the FDR procedures are used to examine between-year differences in subgroup results by jurisdiction. In those cases, results were not included in this report.

---

1. Miller, R. G. (1981). *Simultaneous Statistical Inference* (2nd ed.). New York: Springer-Verlag.

2. Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society,* Series B, No. 1, pp. 289–300.

3. Williams, V. S. L., Jones, L. V., and Tukey, J. W. (1999). Controlling Error in Multiple Comparisons with Examples From State-to-State Differences in Educational Achievement. *Journal of Educational and Behavioral Statistics, 24*(1), 42–69.

4. The level of confidence times the number of comparisons minus one divided by the number of comparisons, or 0.05 x (5-1)/5 = 0.04 = 4 percent.

Understanding NAEP Reporting Groups

NAEP results are provided for groups of students defined by shared characteristics–gender, race/ethnicity, school's type of location, and eligibility for free/reduced-price school lunch. Based on participation rate criteria, results are reported for subpopulations only when sufficient numbers of students and adequate school representation are present. The minimum requirement is at least 62 students in a particular subgroup from at least five primary sampling units (PSUs).[1] However, the data for all students, regardless of whether their subgroup was reported separately, were included in computing overall results. Definitions of the subpopulations are presented below.

Gender: Results are reported separately for males and females.

Race/Ethnicity: In all NAEP assessments, data about student race/ethnicity is collected from two sources: school records and student self-reports. Prior to 2002, NAEP used students' self-reported race as the primary race/ethnicity reporting variable. Starting in 2002, the race/ethnicity variable presented in NAEP reports is based on the race reported by the school. When school-recorded information is missing, student-reported data are used to determine race/ethnicity. The mutually exclusive racial/ethnic categories are White, Black, Hispanic, Asian/Pacific Islander, American Indian (including Alaska Native), and Other. Information based on student self-reported race/ethnicity is available on the NAEP Data Tool (http://nces.ed.gov/nationsreportcard/naepdata/).

Type of Location: Results from the 2003 assessment are reported for students attending schools in three mutually exclusive location types: central city, urban fringe/ large town, and rural/ small town.

Central city: Following standard definitions established by the Federal Office of Management and Budget, the U.S. Census Bureau (see http://www.census.gov/) defines "central city" as the largest city of a Metropolitan Statistical Area (MSA) or a Consolidated Metropolitan Statistical Area (CMSA). Typically, an MSA contains a city with a population of at least 50,000 and includes its adjacent areas. An MSA becomes a CMSA if it meets the requirements to qualify as a metropolitan statistical area, has a population of 1,000,000 or more, its component parts are recognized as primary metropolitan statistical areas, and local opinion favors the designation. In the NCES Common Core of Data (CCD) locale codes are assigned to schools. For the definition of central city used in this report, two locale codes of the survey are combined. The definition of each school's type of location is determined by the size of the place where the school is located and whether or not it is in an MSA or CMSA. School locale codes are assigned by the U.S. Census Bureau. For the definition of central city, NAEP reporting uses data from two CCD locale codes: large city (a central city of an MSA or CMSA with the city having a population greater than or equal to 25,000) and midsize city (a central city of an MSA or CMSA having a population less than 25,000). Central city is a geographical term and is not synonymous with "inner city."

Urban fringe/large town: The urban fringe category includes any incorporated place, census designated place, or non-place territory within a CMSA or MSA of a large or mid-sized city and defined as urban by the U.S. Census Bureau, but which does not qualify as a central city. A large town is defined as a place outside a CMSA or MSA with a population greater than or equal to 25,000.

Rural/small town: Rural includes all places and areas with populations of less than 2,500 that are classified as rural by the U.S. Census Bureau. A small town is defined as a place outside a CMSA or MSA with a population of less than 25,000, but greater than or equal to 2,500.

Results for each type of location are only compared across years 2000 and after. This is due to new methods used by NCES to identify the type of location assigned to each school in the Common Core of Data (CCD). The new methods were put into place by NCES in order to improve the quality of the assignments, and they take into account more information about the exact physical location of the school. The variable was revised in NAEP beginning with the 2000 assessments.

Eligibility for Free/Reduced-Price School Lunch: As part of the Department of Agriculture's National School Lunch Program, schools can receive cash subsidies and donated commodities in turn for offering free or reduced-price lunches to eligible children. Based on available school records, students were classified as either currently eligible for free/reduced-price school lunch or not eligible. Eligibility for the program is determined by students' family income in relation to the federally established poverty level. Free lunch qualification is set at 130 percent of the poverty level, and reduced-price lunch qualification is set at 170 percent of the poverty level. Additional information on eligibility may be found at the Department of Agriculture web site (http://www.fns.usda.gov/cnd/lunch/). The classification applies only to the school year when the assessment was administered (i.e., the 2002–03 school year) and is not based on eligibility in previous years. If school records were not available, the student was classified as "Information not available." If the school did not participate in the program, all students in that school were classified as "Information not available."

---

1. For the national NAEP assessments prior to 2002, a PSU is a selected geographic region (a county, group of counties, or metropolitan statistical area). Since 2002, the first-stage sampling units are schools (public and nonpublic) in the selection of the combined sample.

Where to Find More Information

The NAEP Mathematics Assessment
The latest news about the NAEP 2003 Mathematics assessment and the national results can be found on the NAEP web site at http://nces.ed.gov/nationsreportcard/mathematics/results/. The individual snapshot reports for each participating state and other jurisdictions are also available in the state results section of the web site at http://nces.ed.gov/nationsreportcard/states/. *The Nation's Report Card Mathematics Highlights 2003* may be ordered or downloaded at the NAEP web site. *The Nation's Report Card: Mathematics 2003* will be available at the NAEP web site in 2004. *The Mathematics Framework for the 2003 National Assessment of Educational Progress,* on which this assessment is based, is available at the Internet address http://www.nagb.org/pubs/math_fw_03.pdf.

Additional Results from the Mathematics Assessment
For more findings from the 2003 Mathematics assessments, refer to the NAEP 2003 results at http://nces.ed.gov/nationsreportcard/naepdata/. The interactive database at this site includes student, teacher, and school variables for all participating states and other jurisdictions, the nation, and the four regions. Data tables are also available for each jurisdiction, with all background questions cross-tabulated with the major demographic variables.

Technical Documentation
For explanations of NAEP survey procedures see Allen, N. L., Donoghue, J. R., and Schoeps, T. L. (2001). *The NAEP 1998 Technical Report*. (NCES 2001–509). Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics. Technical information may also be found on the NAEP web site (http://nces.ed.gov/nationsreportcard/mathematics/results2003/interpret-results.asp).

Publications on the inclusion of students with disabilities and limited English proficient students
Olson, J. F., and Goldstein, A. A. (1997). *The Inclusion of Students with Disabilities and Limited-English-Proficient Students in Large-Scale Assessments: A Summary of Recent Progress* (NCES 97–482). Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics.

Mazzeo, J., Carlson, J. E., Voelkl, K. E., and Lutkus, A. D. (2000). *Increasing the Participation of Special-Needs Students in NAEP: A Report on 1998 Research Activities* (NCES 2000–473). Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics.

Lutkus, A. D., and Mazzeo, J. (2003). *Including Special-Needs Students in the NAEP 1998 Mathematics Assessment, Part I: Comparison of Overall Results With and Without Accommodations* (NCES 2003–467). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics.

Lutkus, A. D. (forthcoming). *Including Special-Needs Students in the NAEP 1998 Mathematics Assessment, Part II: Results for Students with Disabilities and Limited English Proficient Students* (NCES 2003–468). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics.

To Order Publications
Recent NAEP publications related to mathematics are listed on the mathematics page of the NAEP web site and are available electronically. Publications can also be ordered from:

Education Publications Center (ED Pubs)
U.S. Department of Education
P.O. Box 1398
Jessup, MD 20794–1398

Call toll free: 1-877-4ED PUBS (1-877-433-7827)
TTY/TDD: 1-877-576-7734
FAX: 1-301-470-1244

The NAEP State Report Generator was developed for the NAEP 2003 reports by Phillip Leung, Jilei Yin, Julian Rosse, Paul Gazzillo, Mike Narcowich, Nancy Mead, Anthony Lutkus, Forton Wimbush, Arlene Weiner, and Patricia Hamill.

What is The Nation's Report Card?

THE NATION'S REPORT CARD, the National Assessment of Educational Progress (NAEP), is a nationally representative and continuing assessment of what America's students know and can do in various subject areas. Since 1969, assessments have been conducted periodically in reading, mathematics, science, writing, history, geography, and other fields. By making objective information on student performance available to policymakers at the national, state, and local levels, NAEP is an integral part of our nation's evaluation of the condition and progress of education. Only information related to academic achievement is collected under this program. NAEP guarantees the privacy of individual students and their families.

NAEP is a congressionally mandated project of the National Center for Education Statistics, within the Institute of Education Sciences of the U.S. Department of Education. The Commissioner of Education Statistics is responsible, by law, for carrying out the NAEP project through competitive awards to qualified organizations.

In 1988, Congress established the National Assessment Governing Board (NAGB) to oversee and set policy for NAEP. The Board is responsible for: selecting the subject areas to be assessed; setting appropriate student achievement levels; developing assessment objectives and test specifications; developing a process for the review of the assessment; designing the assessment methodology; developing guidelines for reporting and disseminating NAEP results; developing standards and procedures for interstate, regional, and national comparisons; determining the appropriateness of all assessment items and ensuring the assessment items are free from bias and are secular, neutral, and nonideological; taking actions to improve the form, content, use, and reporting of results of the National Assessment; and planning and executing the initial public release of National Assessment of Educational Progress reports.

The National Assessment Governing Board